Vol. 47, Parts 3 and 4 December 1960

UN
OF M
JAN 3
SCIENCE
H
81
3615
PR

BIOMETRIKA

FOUNDED BY

W. F. R. WELDON, FRANCIS GALTON AND KARL PEARSON

MANAGING EDITOR

E. S. PEARSON

ASSOCIATE EDITOR

M. G. KENDALL

ISSUED BY

THE BIOMETRIKA OFFICE, UNIVERSITY COLLEGE LONDON

[*Issued December* 1960]

This volume of *Biometrika* is published with the co-operation of

F. N. DAVID	N. L. JOHNSON
J. DURBIN	D. V. LINDLEY
M. J. R. HEALY	R. L. PLACKETT

A volume containing about 500 pages will be published annually in two half-yearly issues, appearing in June and December.

Papers for publication should be sent to

PROFESSOR E. S. PEARSON

University College London, Gower Street, London, W.C. 1

It is a condition of publication in *Biometrika* that the paper shall not already have been issued elsewhere, and will not be reprinted without leave of the Editors.

Contributors receive 50 copies of their papers free. Joint authors 25 copies each. Order forms for separates are sent to authors with proofs of their papers.

SUBSCRIPTIONS for 1961 (Volume 48)

The subscription price, payable in advance is:
£2. 14*s.* (or **$8.00**) net per volume *including packing and postage.*

Subscribers who buy the journal *Biometrika* for their own and not for general use may apply to the Secretary to be registered as **Personal Subscribers,** for whom the subscription price is **£2. 5*s.*** (payable only in sterling).

Cheques should be made payable to *Biometrika,* crossed 'a/c BIOMETRIKA TRUST' and sent to THE SECRETARY, BIOMETRIKA OFFICE, UNIVERSITY COLLEGE LONDON, GOWER STREET, LONDON, W.C. 1, to whom all orders for Offprints, Index and Table Separates (see pages (iv) to (vi)) should also be sent. For particulars as to the availability of **Back Issues** please see inside back cover.

[Members of the Institute of Mathematical Statistics are reminded that they may subscribe to *Biometrika* through the Institute who will accept subscriptions to *Biometrika* at the special rate of $6.50 if paid to the Treasurer of the Institute before 1 March.]

Biometrika (1960), **47**, 3 *and* 4, *p.* 219
With 2 text-figures
Printed in Great Britain

The properties of a stochastic model for the predator-prey type of interaction between two species

BY P. H. LESLIE AND J. C. GOWER

Bureau of Animal Population, Department of Zoological Field Studies, Oxford, and the Statistical Department, Rothamsted Experimental Station

1. INTRODUCTION

The two main types of interaction between any pair of biological species, which are of most interest to the ecologist, are either when they are competing together for some common source of food supply, or when one of the species preys upon the other. The properties of a stochastic model for a system of two competing species have been illustrated by a set of artificial realizations in an earlier paper (Leslie & Gower, 1958). We now consider some of the properties of a model for the interaction between a predatory species and its prey, more particularly in regard to the behaviour of such systems in the region of the stationary state, and the chances of random extinction for one or other of the two species. The interaction is imagined to take place in some limited environment where there is an ample supply of food for the prey, and in the first place it is assumed that all the members of the prey population are exposed to the risk of capture or attack by the predator. Later, this last condition is relaxed, and it is supposed that because of some heterogeneity in the environment, only a fraction of the prey population is exposed to risk, and some of the consequences of this partial interaction are illustrated in the last section. The basic model has already been described (Leslie, 1948, 1958), but it will be useful if the formulae are summarized here.

If S_1 is a species of prey and S_2 a predator, then working in discrete-time intervals and given that the respective populations consist of $N_1(t)$ and $N_2(t)$ individuals at time t, the deterministic model for the interaction is taken to be,

$$\left.\begin{aligned} N_1(t+1) &= \frac{\lambda_1 N_1(t)}{1+\alpha_1 N_1(t)+\beta_1 N_2(t)}, \\ N_2(t+1) &= \frac{\lambda_2 N_2(t)}{1+\alpha_2\{N_2(t)/N_1(t)\}}, \end{aligned}\right\} \tag{1·1}$$

where α_1, β_1 and α_2 are parameters > 0, and $\log_e \lambda_1 = r_1 = b_1 - d_1$, $\log_e \lambda_2 = r_2 = b_2 - d_2$ are the intrinsic birth- and death-rates of the two species. For the stationary state, $N_1 = L_1$, $N_2 = L_2$, we have,

$$\left.\begin{aligned} L_1 &= \frac{\alpha_2(\lambda_1 - 1)}{\beta_1(\lambda_2 - 1)+\alpha_1\alpha_2}, \\ L_2 &= (\lambda_2 - 1)\, L_1/\alpha_2. \end{aligned}\right\} \tag{1·2}$$

It has been shown (Leslie, 1948, § 6) that this discrete-time model is closely related to the continuous-time model, (writing the operator $D \equiv d/dt$),

$$\left.\begin{aligned} DN_1 &= (r_1 - a_1 N_1 - c_1 N_2)\, N_1 \\ DN_2 &= \left(r_2 - a_2 \frac{N_2}{N_1}\right) N_2 \end{aligned}\right\} \quad (a_1, a_2, c_1 > 0), \tag{1·3}$$

for which the stationary state is

$$\left.\begin{aligned} L_1 &= \frac{a_2 r_1}{c_1 r_2 + a_1 a_2}, \\ L_2 &= r_2 L_1 / a_2. \end{aligned}\right\} \tag{1·4}$$

Putting $N_1 = L_1(1+u_1)$ and $N_2 = L_2(1+u_2)$, then we have from (1·3) for small u_1 and u_2,

$$\left.\begin{aligned} Du_1 &= -a_1 L_1 u_1 - c_1 L_2 u_2, \\ Du_2 &= r_2(u_1 - u_2). \end{aligned}\right\} \tag{1·5}$$

The equation in D is $$D^2 + (r_2 + a_1 L_1) D + r_2(a_1 L_1 + c_1 L_2) = 0,$$

or from (1·4) $$D^2 + (r_2 + a_1 L_1) D + r_1 r_2 = 0. \tag{1·6}$$

The roots of (1·6), therefore, will be either both negative, or complex with the real part negative, so that the stationary state is stable, and an approach to this state will be made by a series of damped oscillations if

$$(r_2 + a_1 L_1)^2 < 4 r_1 r_2.$$

The properties of this model differ somewhat from those of the classical Lotka–Volterra equations for this type of interaction, viz.

$$\left.\begin{aligned} DN_1 &= (\epsilon_1 - \gamma_1 N_2) N_1 \\ DN_2 &= (-\epsilon_2 + \gamma_2 N_1) N_2 \end{aligned}\right\} \quad (\epsilon_1, \epsilon_2, \gamma_1, \gamma_2 > 0). \tag{1·7}$$

This deterministic system and the stochastic model based upon it have been discussed recently by Bartlett (1957*a*). Here there is no damping towards the equilibrium point $L_1 = \epsilon_2/\gamma_2$, $L_2 = \epsilon_1/\gamma_1$, and the deterministic path of small cycles around this point is given by the ellipse

$$\epsilon_2 u_1^2 + \epsilon_1 u_2^2 = \text{constant},$$

where $N_1 = L_1(1+u_1)$, $N_2 = L_2(1+u_2)$. Consequently, as Bartlett points out, in the stochastic model the effect of random drift, in the absence of any damping factor, must result sooner or later, either in the extinction of the predator, or in the extinction first of the prey, followed in due course by that of the predator from starvation. In the stochastic formulation of this model (1·7), the stationary state is therefore in a sense unstable. But the presence of a damping term in the model defined by (1·3) would lead to a greater degree of stability, and in the corresponding stochastic model the system might therefore be expected to fluctuate in the region of the equilibrium state, although there might be still a chance of random extinction of one or other species, depending on the degree of variability in numbers in this region. To investigate this point, the stochastic discrete-time model corresponding to (1·1) was programmed for the Elliot–N.R.D.C. computer in the Statistical Department of Rothamsted Experimental Station, and some of the results obtained are given in the following sections.

2. Stochastic model

In a stochastic formulation of this model, a wide variety of possible birth-rate and death-rate functions for the two species could be adopted (cf. Leslie, 1958, § 4). But, clearly, one of the most obvious choices to make is the case when the birth-rate (b) of each species remains constant, and any increase in the numbers of the predator affects the death-rate ($d_1(t)$) of the prey over the interval t to $t+1$, and vice versa. If, therefore, in (1·1) we define the parameters representing the intrinsic birth- and death-rates of the two species,

$$\lambda_a = e^{b_a - d_a} \quad (a = 1, 2;\ b_a > d_a),$$

and put

$$q_1(t) = 1 + \alpha_1 N_1(t) + \beta_1 N_2(t),$$
$$q_2(t) = 1 + \alpha_2\{N_2(t)/N_1(t)\},$$

we have in the constant birth-rate type of discrete-time stochastic model (Leslie, 1958, § 5),

$$\left.\begin{aligned} E\{N_a(t+1) \mid N_1(t), N_2(t)\} &= \frac{\lambda_a}{q_a(t)} N_a(t) = \lambda_a(t)\, N_a(t) \quad (a = 1, 2), \\ \operatorname{var}\{N_a(t+1) \mid N_1(t), N_2(t)\} &= \left(\frac{2b_a}{r_a(t)} - 1\right)(\lambda_a(t) - 1)\, E\{N_a(t+1)\} \quad (\lambda_a(t) \neq 1) \\ &= 2b_a E\{N_a(t+1)\} \quad (\lambda_a(t) = 1). \end{aligned}\right\} \tag{2·1}$$

Here it is assumed that the difference between the constant birth-rate and the variable death-rate of each species at time t,

$$r_a(t) = \log_e[\lambda_a/q_a(t)] = b_a - d_a(t) \quad (a = 1, 2)$$

remains constant over the interval t to $t+1$, and that given $N_1(t)$ and $N_2(t)$, then $N_1(t+1)$ and $N_2(t+1)$ are distributed normally and independently with these means and variances, all negative values of $N_a(t+1)$ being attributed to $N_a(t+1) = 0$.

The programmes for this model were similar to those described by Leslie & Gower (1958) for a system of two competing species and little need be said of them here. In particular the same method of random number generation was used and the population parameters could easily be varied. Since the results consisted largely of long trajectories of the population numbers of the two species at succeeding instants of time, an additional programme was written to construct distributions of these population numbers direct from the output tapes.

3. Theoretical variances and covariance of the system

The theoretical variances and covariance of the population numbers, when such a system is fluctuating in the region of the stationary state, may be estimated, to a first approximation, in the following way. The equations (2·1) may be written in the form

$$N_a(t+1) = f_a(N_1, N_2, t) + Z_a(t+1) \quad (a = 1, 2) \tag{3·1}$$

in which the first term is the deterministic part of the process defined by (1·1); and the $Z_a(t+1)$ are independent normal variables with zero means, and variances, $\sigma^2(Z_a) \sim 2b_a L_a$, when the birth-rate of each species remains constant, and the system is in the region of the stationary state (L_1, L_2) defined by (1·2). By applying the usual δ-technique (cf. Leslie & Gower, 1958, § 5), we have as a first approximation from (3·1), the set of equations for determining $\operatorname{var}(N_1(t)) = \sigma_1^2$, $\operatorname{var}(N_2(t)) = \sigma_2^2$ and $\operatorname{cov}(N_1(t), N_2(t)) = \sigma_{12}$,

$$\left.\begin{aligned} \sigma_1^2 &= \left(\frac{\partial f_1}{\partial N_1}\right)^2 \sigma_1^2 + 2\left(\frac{\partial f_1}{\partial N_1}\right)\left(\frac{\partial f_1}{\partial N_2}\right)\sigma_{12} + \left(\frac{\partial f_1}{\partial N_2}\right)^2 \sigma_2^2 + 2b_1 L_1, \\ \sigma_2^2 &= \left(\frac{\partial f_2}{\partial N_1}\right)^2 \sigma_1^2 + 2\left(\frac{\partial f_2}{\partial N_1}\right)\left(\frac{\partial f_2}{\partial N_2}\right)\sigma_{12} + \left(\frac{\partial f_2}{\partial N_2}\right)\sigma_2^2 + 2b_2 L_2, \\ \sigma_{12} &= \left(\frac{\partial f_1}{\partial N_1}\right)\left(\frac{\partial f_2}{\partial N_1}\right)\sigma_1^2 + \left[\left(\frac{\partial f_1}{\partial N_1}\right)\left(\frac{\partial f_2}{\partial N_2}\right) + \left(\frac{\partial f_1}{\partial N_2}\right)\left(\frac{\partial f_2}{\partial N_1}\right)\right]\sigma_{12} + \left(\frac{\partial f_1}{\partial N_2}\right)\left(\frac{\partial f_2}{\partial N_2}\right)\sigma_2^2, \end{aligned}\right\} \tag{3·2}$$

the partial differential coefficients being evaluated for $N_1 = L_1$, $N_2 = L_2$. Thus, from (1·1) and (1·2), we have

$$f_1 = \frac{\lambda_1 N_1}{1+\alpha_1 N_1+\beta_1 N_2},$$

$$f_2 = \frac{\lambda_2 N_2}{1+\alpha_2 N_2/N_1},$$

and

$$\left(\frac{\partial f_1}{\partial N_1}\right)_{L_1 L_2} = 1-\frac{\alpha_1 L_1}{\lambda_1}, \quad \left(\frac{\partial f_2}{\partial N_2}\right)_{L_1 L_2} = \frac{1}{\lambda_2},$$

$$\left(\frac{\partial f_1}{\partial N_2}\right)_{L_1 L_2} = -\frac{\beta_1 L_1}{\lambda_1}, \quad \left(\frac{\partial f_2}{\partial N_1}\right)_{L_1 L_2} = \frac{(\lambda_2-1)^2}{\alpha_2 \lambda_2}.$$

It is also useful to have the equations for determining the variances and covariance for the analogous continuous-time model. From (1·5), writing $c_1 = ka_1$, we have the linear stochastic system for small u_1 and u_2,

$$\left.\begin{aligned} \delta u_1 &= -a_1(L_1 u_1 + kL_2 u_2)\,\delta t + \delta\xi, \\ \delta u_2 &= r_2(u_1 - u_2)\,\delta t + \delta\phi, \end{aligned}\right\} \tag{3·3}$$

in which the independent variables $\delta\xi$ and $\delta\phi$ have zero means, and variances $(2b_1/L_1)\,\delta t$ and $(2b_2/L_2)\,\delta t$, respectively, when the birth-rate of each species is assumed to remain constant and the system is in the region of the state (L_1, L_2). From $u_1+\delta u_1$, $u_2+\delta u_2$, we finally obtain, by squaring and taking the cross-product, the equations for the variances and covariance $(\sigma_1^2, \sigma_2^2, \sigma_{12})$ of u_1 and u_2,

$$\left.\begin{aligned} L_1\sigma_1^2 &= \frac{b_1}{a_1 L_1} - kL_2\sigma_{12}, \\ \sigma_2^2 &= \frac{b_2}{r_2 L_2} + \sigma_{12}, \\ (a_1 L_1 + r_2)\,\sigma_{12} &= r_2\sigma_1^2 - ka_1 L_2\sigma_2^2. \end{aligned}\right\} \tag{3·4}$$

Since $N_1 = L_1(1+u_1)$, $N_2 = L_2(1+u_2)$, we then have

$$\operatorname{var}(N_1) = L_1^2 \operatorname{var}(u_1), \quad \operatorname{var}(N_2) = L_2^2 \operatorname{var}(u_2), \quad \operatorname{cov}(N_1, N_2) = L_1 L_2 \operatorname{cov}(u_1, u_2),$$

which can be contrasted with the variances and covariance for the discrete-time model given by (3·2). The relationships between the parameters of the two types of model are

$$r_a = \log_e \lambda_a \quad (a = 1, 2), \qquad a_1 = \alpha_1 r_1/(\lambda_1 - 1) \qquad \text{and} \qquad k = \beta_1/\alpha_1;$$

while the constant birth-rates b_1 and b_2, and the stationary values, L_1 and L_2, are naturally the same for each model. In regard to the differences between the two models, it is perhaps worth noting that the variances estimated by means of (3·4) are always less than those obtained by the solution of (3·2), so that the use of the discrete-time model leads to an enhanced degree of variation when a particular system is fluctuating in the neighbourhood of the equilibrium level (cf. Bartlett, Gower & Leslie, 1960). By reducing the unit of time adopted in the discrete-time model, however, it is always possible to make the variances of the two types of model more in agreement.

4. Fluctuations in the neighbourhood of the stationary state

The behaviour of this type of system in the region of the stable stationary state may be illustrated by the results obtained from a realization which was computed for 1000 steps, taking as the numerical values of the parameters in (2·1)

$$\lambda_1 = 1{\cdot}2574, \quad b_1 = 0{\cdot}2577, \quad \lambda_2 = 1{\cdot}1892, \quad b_2 = 0{\cdot}2521,$$
$$\alpha_1 = 0{\cdot}0005148, \quad \beta_1 = 0{\cdot}0018018 \quad \text{and} \quad \alpha_2 = 0{\cdot}2838,$$

which from (1·2) give a stationary state of $L_1 = 150$, $L_2 = 100$. Starting with the initial values of $N_1(0) = 500$ for the prey, and $N_2(0) = 20$ for the predator, these processes rapidly approached their equilibrium levels, in the region of which they then continued to oscillate

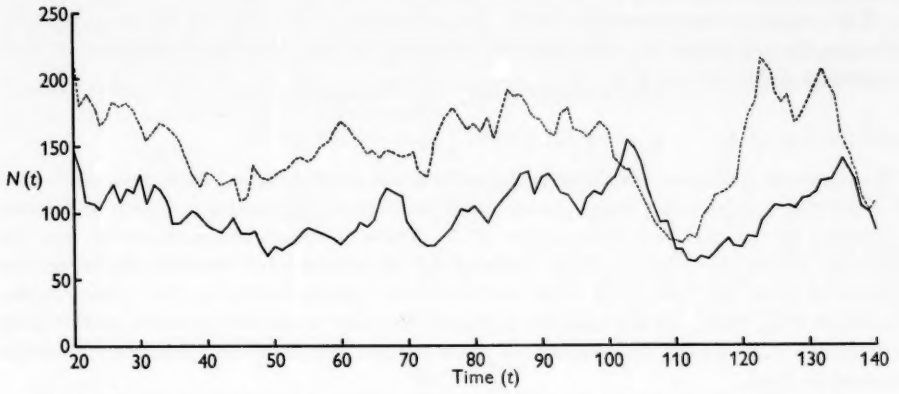


Fig. 1. Fluctuations of a predator–prey system in the region of the stable stationary state. – – –, the number of prey (equilibrium level $L_1 = 150$); ——, the number of predators (equilibrium level $L_2 = 100$).

in an irregular fashion. As an example, the changes in numbers are illustrated in Fig. 1 for the period of time from $t = 20$ to $t = 140$. It will be seen that there is no regularity in the oscillations, and that at times there is a tendency for quite a pronounced excursion to be made away from the region of the stationary state, for example, in Fig. 1 the course of events from about $t = 100$ to $t = 120$, when a large fall in the numbers of the prey is followed shortly afterwards by a similar type of fall in numbers of the predator. The general picture is somewhat different from that of a system of two competing species with a stable stationary state (Leslie & Gower, 1958), where any drift away from the equilibrium level seems to take place in opposite directions owing to the negative correlation which exists between the two species in that type of system. In the case of the predator–prey type of interaction, the correlation is positive in sign.

Theoretically, it might be expected that the stationary distribution of such a two-species system would be approximately bi-variate normal in form, with variances and covariance given by (3·2), (Bartlett, Gower & Leslie, 1960). The bivariate distribution of $N_1(t)$ and $N_2(t)$ is given in Appendix Table 1 for the last 800 steps in this realization. The observed means were $\overline{N}_1 = 149{\cdot}74$ and $\overline{N}_2 = 97{\cdot}09$, compared with the deterministic stable state of $L_1 = 150$, $L_2 = 100$; while the observed variances and covariance (computed by the machine from the

individual observations and not from the grouped data), together with the values expected theoretically both for the discrete-time model, and for its continuous-time equivalent, were:

	Expected		Observed
	Continuous-time model	Discrete-time model	Discrete-time model
$\text{var}(N_1)$	432	546	635·6
$\text{var}(N_2)$	170	228	214·0
$\text{cov}(N_1, N_2)$	+37·2	+56·6	+63·6

Considering that the expected values are based on the theory of small fluctuations and so cannot be exactly correct, the agreement in the case of the model actually used seems very satisfactory.

For samples as large as 800, taking $g_1 \sim m_3/m_2^{\frac{3}{2}}$ and $g_2 \sim [(m_4/m_2^2) - 3]$, we have (applying Sheppard's corrections in computing the moments) for the marginal distribution of N_1 in Appendix Table 1,

$$g_1 \sim +0{\cdot}322 \pm 0{\cdot}087, \quad g_2 \sim -0{\cdot}035 \pm 0{\cdot}173,$$

and for that of N_2,

$$g_1 \sim +0{\cdot}121 \pm 0{\cdot}087, \quad g_2 \sim -0{\cdot}201 \pm 0{\cdot}173.$$

The skewness coefficient for the second species is not much different from zero, while that for the first species is 3·7 times the standard error given, suggesting a degree of positive skewness in the marginal distribution of N_1. These standard errors, however, are the classical values, viz. $\sigma(g_1) = \sqrt{(6/n)}$, ignoring the serial correlation between the successive values of both $N_1(t)$ and $N_2(t)$. They are therefore underestimates of the correct values (Bartlett *et al.* 1960). In any case the degree of skewness is not very marked, and broadly speaking, we can regard the distribution given in Table 1 as being approximately bivariate normal in form.

5. The chances of random extinction

In the numerical system used as an illustration in the previous section, the chance of one or other of the species becoming extinct appears to be negligible, at least in terms of any ordinary time scale, once their numbers are in the region of their equilibrium levels. For instance, to take the predatory species which has the smaller numbers as an example; the order of magnitude of the mean passage time to extinction is likely to be inversely proportional to the probability of the zero state, or in terms of a normal distribution, using the observed mean and variance for this species, of the order of $1{\cdot}4 \times 10^{11}$ units of time. It might be expected, therefore, that this system would continue to oscillate around the equilibrium level in an irregular fashion for a comparatively lengthy period. But, in other systems, this may not be so, and the chances of extinction may then become appreciable.

This point is relevant to the results obtained by Gause (1934) in his experiments with the two Protozoa, *Didinium nasutum* and *Paramecium caudatum*. The latter species is devoured by the former, and the amount of food required by *Didinium* is very great. His first experiments were carried out in a relatively small microcosm (the upper asymptote in numbers of the *Paramecia*, when living alone, being of the order of 100 individuals), and the introduction of small numbers of *Didinium* at varying intervals of time after the prey was followed by an intense multiplication of the predator, which devoured all the prey and afterwards themselves perished. There was also a considerable degree of variability in numbers between replicates of the same age (Gause, 1934, Chapter VI and Appendix, Table 6).

Suppose, then, we have a relatively small environment in which a species of prey is established, and which is then invaded by a small group of predators. For simplicity it will be assumed in the model that the intrinsic rates of increase and the birth-rates of both species are the same, and that in (2·1) we put

$$\lambda_1 = \lambda_2 = 1{\cdot}25, \quad b_1 = b_2 = 0{\cdot}25, \quad \alpha_1 = 0{\cdot}001.$$

These values of λ_1 and α_1 lead to a logistic upper-asymptote in numbers of the prey, if it were living alone, of 250 individuals. When it is assumed that the birth-rates of both species remain constant, the remaining parameters, β_1 and α_2, in (1·1) and (2·1), define the magnitude of the effect which the numbers of each species have on the death-rate of the other. β_1 might be termed the 'relative voracity' of the predator in regard to the prey, since for a given value of N_2, the greater the value of this parameter, the more the death-rate of the prey is affected during a unit of time; and α_2 represents the sensitivity of the predator to the relative density in numbers of the two species, i.e. for a given value of the ratio N_2/N_1, the greater the value of α_2, the more the death-rate of the predator is affected. Consider three systems in which β_1 and α_2 are varied slightly, which together with the values of the other parameters given above lead to the following stationary states:

System	β_1	α_2	L_1	L_2
(i)	0·015	0·375	22·7	15·2
(ii)	0·025	0·500	18·5	9·3
(iii)	0·050	0·250	4·9	4·9

Some realizations were computed for each of these three systems, taking in each case $N_1(0) = 200$, $N_2(0) = 15$ as the initial conditions, i.e. it was assumed that the prey was well established in this imaginary environment, when it was invaded by a small group of predators. The results were as follows.

System (i)

The computations were carried on for 399 steps, at the end of which time neither species was eliminated, though at one point the predator was very close to extinction, having been reduced to only one individual surviving on two successive occasions. Neglecting the first ten steps, during which time an approach was being made to the region of the stationary state, the observed means for the remaining 389 observations were $\overline{N}_1 = 25{\cdot}33$ and $\overline{N}_2 = 12{\cdot}96$; while the observed variances and covariance compared with the theoretical expectations derived from (3·3) were

	Expected	Observed
var (N_1)	101	137·3
var (N_2)	37·7	33·35
cov (N_1, N_2)	+19·7	+22·02

This very good agreement between the theoretical and observed variances was of interest, since in this type of discrete-time stochastic model, the assumption made in applying (2·1) that the $N_a(t+1)$ $(a = 1, 2)$, for given $N_a(t)$ are distributed normally, might be expected to break down in the region of small numbers. The observed distributions of N_1 and N_2 are given in Appendix Table 2, from which it can be seen that these departed appreciably from the normal, there being a pronounced positive skewness in each case. However, with no knowledge of the form of distribution which might be expected in such systems when numbers are small, it is difficult to know what reliance can be placed on this observed bivariate distributic

System (ii)

The predator became extinct in each of the five realizations which were computed for this system. In one case the prey also vanished at the same time as the predator, and in another its numbers were reduced to only one individual. In this region of a stochastic logistic process, the probability of random extinction should be roughly of the same order of magnitude as that of a simple birth and death process with parameters corresponding to the intrinsic rate of increase of the particular species, viz. $\log \lambda = b - d$; so that if only one individual is left surviving, the probability that the prey population would in turn become extinct, is appreciable in this numerical example, being $P(0) \sim d/b = 0{\cdot}11$. In other words, the disappearance of the predator before the prey may not necessarily mean that the latter species will be able to persist.

System (iii)

Out of 19 realizations with the initial numbers $N_1(0) = 200$, $N_2(0) = 15$, the prey was eliminated by the predator in 9 cases, which would result inevitably in the disappearance also of the predator. In the remaining 10, the predator became extinct before the prey; but

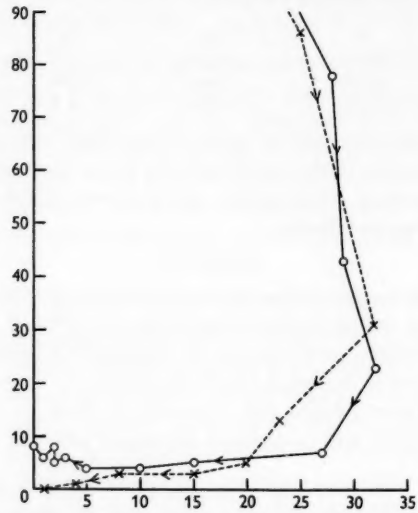


Fig. 2. Representative trajectories for System (iii), for which the deterministic stable state is $L_1 = L_2 \sim 5$. The initial conditions were for the prey, $N_1(0) = 200$, and for the predator, $N_2(0) = 15$, but in order to save space, a few of the initial steps in each realization are omitted from the figure.

in five of these cases the prey at this time was reduced to such low numbers, for example, either one or two individuals, that its ability to persist became problematical. There appeared to be little difference between the times taken for either predator or prey to become extinct, the combined mean passage time to extinction being $\bar{t} = 20{\cdot}3 \pm 2{\cdot}99$ units of time. The (N_1, N_2) trajectories differed somewhat, depending on the length of time the interaction persisted, but a pair of replicates are illustrated in Fig. 2, for which the passage time to zero was of the same order as the mean value. In broad outline the pattern is very similar to the trajectories observed by Gause (1934) in his experiments, not only with *Didinium*

and *Paramecium*, but also with other species of Protozoa, for example, *Paramecium bursaria* and *Bursaria truncatella* (Gause, 1935).

The results with these three numerical systems suggest that, in a limited environment, in which there is no immigration from outside, the persistence of both species in a predator–prey interaction depends very greatly on the relations between the various parameters for the two species, and that the chances of random extinction are very sensitive to relatively small changes in these parameters. If the magnitudes of the intrinsic rates of increase of the two species, the relative 'voracity' of the predator, and so forth, are not happily adjusted in between themselves, it appears doubtful whether in a very simplified type of system both species will be able to persist for very long, and the probability of random extinction then becomes a dominant feature of the interaction.

6. When only a fraction of the prey population is exposed to risk

In developing the type of model we have used so far, one tacit assumption has been made, namely, that in some way or other the entire prey population is exposed to the risk of capture by the predator. The background environment in which the interaction takes place is assumed to be in a sense 'homogeneous', and the predator and prey populations freely intermingle. Suppose, however, that this is not so, and that because of some 'heterogeneity' of the environment, only a fraction of the prey population is exposed to risk during any one unit of time. One way in which this might occur would be if the configuration of some type of vegetation formed a 'cover' for the prey, which protected it from the predator. Thus, Gause (1934) has pointed out the importance of a refuge in enabling the prey to survive in his experiments with various species of Protozoa, although the predator might still become extinct. But, apart from 'cover' in its literal sense, a variety of different situations can be imagined as regards the environment, or the habits of the two species, which would result in the interaction being confined to only a fraction of the total prey population.

Suppose, for simplicity, that a fraction k $(0 \leqslant k \leqslant 1)$ of the prey population is exposed to risk of capture by the predator, and that this fraction remains constant for a particular system. Then, in a more generalized model in the place of (1·1) and (2·1), the deterministic part of the process may be written,

$$\left.\begin{aligned} E\{N_1(t+1) \mid N_1(t), N_2(t)\} &= \left[\frac{(1-k)\lambda_1}{1+\alpha_1 N_1(t)} + \frac{k\lambda_1}{1+\alpha_1 N_1(t)+\beta_1 N_2(t)}\right] N_1(t), \\ E\{N_2(t+1) \mid N_1(t), N_2(t)\} &= \frac{\lambda_2 N_2(t)}{1+(\alpha_2/k)\{N_2(t)/N_1(t)\}}. \end{aligned}\right\} \tag{6·1}$$

When $k = 1$, this system reduces to (1·1), and when $k = 0$, i.e. when none of the prey population are exposed to risk, the predator cannot survive, and the system becomes a simple logistic model for the prey population when living alone.

For the stationary state (L_1, L_2) of (6·1), we have for the predatory species

$$\left.\begin{aligned} L_2 &= \frac{k(\lambda_2-1)}{\alpha_2} L_1; \\ \text{while that of the prey is given by the positive root of the equation,} \\ L_1^2 + \left[\frac{1-(1-k)\lambda_1}{\alpha_1} + \frac{1-k\lambda_1}{B}\right] L_1 - \frac{\lambda_1 - 1}{\alpha_1 B} &= 0, \\ \text{where} \qquad B &= \alpha_1 + \frac{k\beta_1(\lambda_2-1)}{\alpha_2}. \end{aligned}\right\} \tag{6·2}$$

The theoretical variances and covariance of the stochastic model for this type of system in the region of the stationary state may be obtained from (3·3), in a similar way to those of the original model, allowing for the alterations in the functions f_1 and f_2.

It is of interest to consider the effect of allowing the value of the constant k to vary in a system with a given set of the remaining parameters in (6·1), more particularly in regard to the chances of random extinction. As an illustration, take the System (iii) of the previous section, for which

$$\lambda_1 = \lambda_2 = 1{\cdot}25, \quad \alpha_1 = 0{\cdot}001, \quad \beta_1 = 0{\cdot}050,$$

$$b_1 = b_2 = 0{\cdot}25, \quad \alpha_2 = 0{\cdot}250.$$

In this case it is known from the realizations already computed that the interaction cannot persist for very long, when $k = 1$. Suppose, now, that k is allowed to vary, and that a particular system with constant k is in the region of the stationary state (L_1, L_2) given by (6·2). If, as an approximation, we assume that $N_1(t)$ and $N_2(t)$ are both distributed normally about means given by this state, then the mean passage time to zero for each species should be, very roughly, of the order of magnitude

$$\Theta_a(0) \sim \sqrt{(2\pi)}\,\sigma_a e^{L_a^2/2\sigma_a^2} \quad (a = 1, 2),$$

where the variance σ_a^2 is obtained from (3·3). The following are the stationary states (L_1, L_2) and the estimated values of $\Theta_a(0)$ for a system with the basic parameters given above, over a range of values of k $(0 \leqslant k \leqslant 1)$.

	Stationary state		Order of magnitude of the mean passage time to zero in units	
k	L_1	L_2	$\Theta_1(0)$	$\Theta_2(0)$
0	250·0	0	$5{\cdot}8 \times 10^{40}$	—
0·05	230·1	11·5	$9{\cdot}8 \times 10^{34}$	$6{\cdot}2 \times 10^{2}$
0·10	194·0	19·4	$6{\cdot}0 \times 10^{25}$	$1{\cdot}3 \times 10^{4}$
0·15	155·7	23·4	$7{\cdot}4 \times 10^{17}$	$4{\cdot}8 \times 10^{4}$
0·20	125·0	25·0	$1{\cdot}2 \times 10^{13}$	$6{\cdot}5 \times 10^{4}$
0·25	90·7	22·7	$1{\cdot}5 \times 10^{8}$	$1{\cdot}7 \times 10^{4}$
0·35	50·4	17·6	$2{\cdot}4 \times 10^{4}$	$1{\cdot}5 \times 10^{3}$
0·50	23·3	11·6	$3{\cdot}3 \times 10^{2}$	$1{\cdot}6 \times 10^{2}$
0·65	12·8	8·3	76	58
0·80	8·1	6·5	37	34
1·00	4·9	4·9	20	21

It will be seen that, as might be expected, the stationary numbers of the prey (L_1) steadily increase as a smaller fraction of the population is exposed to risk; while those of the predator (L_2) steadily increase with decreasing k up to a maximum which, in this particular example, is reached in the region of $k = 0{\cdot}20$. The estimates of $\Theta_a(0)$, which must not be taken too literally, change very rapidly as k decreases. Taking the figures for the predator, which is the species having the smaller numbers, it appears that the approximate maximum mean passage time to extinction is reached at $k \sim 0{\cdot}20$, when it is about 3000 times greater than in the case of the original system with $k = 1$. These estimates of $\Theta_a(0)$ suggest that in this particular system both species would persist for a comparatively long time when some 10–35 % of the prey population were exposed to risk, in contrast with the relatively rapid extinction of one or other species as this fraction approaches unity.

It was of interest, experimentally, to see how far some of these figures could be confirmed in practice. Taking $k = 1$ in the model, 100 realizations were computed, starting each time

with the initial numbers $N_1(0) = N_2(0) = 5$, i.e. the nearest integers to the deterministic stationary state. In 51 of the replicates the predator disappeared before the prey, the mean passage time to the zero state being $\bar{t} = 19{\cdot}61 \pm 2{\cdot}27$ units. In the remaining 49 the prey became extinct before the predator, with $\bar{t} = 19{\cdot}25 \pm 2{\cdot}55$. The distributions of the individual times to extinction were very skew, the longest time recorded being in the $t = 76$–79 class.* But these mean figures are in excellent agreement with those given in the table above. It is also of interest to note that these mean times may be very little affected by the initial numbers adopted in this particular system, as was illustrated in the previous section (System iii), when with $N_1(0) = 200$, $N_2(0) = 15$, the combined mean passage time to extinction for 19 replicates was $t \sim 20$ units.

When $k = 0{\cdot}8$ in this system, the 'normal' approximation suggests that the mean passage time to zero should be of the order of 35 steps for both species. In 50 realizations using this value of k, and taking as the initial state $N_1(0) = 8$, $N_2(0) = 7$, the predator became extinct in 38 cases, with $\bar{t} = 43{\cdot}55 \pm 6{\cdot}36$; while in the remaining 12, the prey disappeared in $\bar{t} = 50{\cdot}17 \pm 11{\cdot}48$ steps. Both of these means are somewhat greater than the expected figure, but the distributions of the individual times were even more skew than in the preceding examples when $k = 1$, the interaction on one occasion persisting for 172–175 steps before the predator became extinct. In general, however, these computed mean times to extinction for the two values of k are in reasonably good agreement, at least as regards orders of magnitude, with those derived from the normal approximation.

A further realization was computed for the case of $k = 0{\cdot}20$, for which, according to the above table, the chance of the predator being able to persist is at a maximum, and starting with the initial numbers of $N_1(0) = 125$, $N_2(0) = 25$, the calculations were carried on for 999 steps. During this time there appeared to be no definite approach to the zero state, although on one occasion the numbers in the predator population fell to only six individuals. The observed means for the 999 observations were $\overline{N}_1 = 120{\cdot}75$, $\overline{N}_2 = 22{\cdot}37$, compared with the deterministic stationary state of $L_1 = 125$, $L_2 = 25$; and the variances and covariance were $\mathrm{var}(N_1) = 345{\cdot}2$, $\mathrm{var}(N_2) = 37{\cdot}66$ and $\mathrm{cov}(N_1, N_2) = +8{\cdot}24$. These observed variances are in very good agreement with the theoretical approximations derived from (3·3), namely $\sigma_1^2 \sim 296$, $\sigma_2^2 \sim 37{\cdot}4$ and $\sigma_{12} \sim +7{\cdot}53$. The grouped frequency distribution of N_1 and N_2 is given in Appendix Table 3, and from the marginal distributions we have for the prey species (N_1)

$$g_1 \sim -0{\cdot}168 \pm 0{\cdot}077, \quad g_2 \sim -0{\cdot}089 \pm 0{\cdot}155,$$

and for the predatory species (N_2)

$$g_1 \sim +0{\cdot}131 \pm 0{\cdot}077, \quad g_2 \sim -0{\cdot}133 \pm 0{\cdot}155,$$

where the standard errors attached to the estimates are again the classical values, representing lower limits to the correct values. The magnitude of the skewness coefficients suggests that the distribution can be regarded as approximately bivariate normal in form. But, the main point of these results is to confirm that in this system, both species should co-exist for comparatively lengthy periods of time when, owing to some 'heterogeneity' of the environment, the fraction of the prey population exposed to risk is of the order

* In carrying out these computations, the machine was instructed to stop whenever the number of one or other species reached zero, and to record the number of steps taken for this point to be reached. Although the means and variances were computed from the individual figures, the frequency distributions of the times to extinction were finally printed in groups of four units, so that it is impossible to give the exact highest and lowest figure observed.

of 20%. Naturally, this optimum fraction, which reduces the chance of random extinction of the predator to a minimum, will vary from system to system and may indeed be insufficient to prevent the disappearance of one or other species.

The 'heterogeneity' of the environment is not, however, the only factor which would tend to increase the probability of both species co-existing. The relative dispersion and movements of the two species over an area may also be important. This has been illustrated very elegantly by some experiments by Huffaker (1958) with the mite *Eotetranychus sexmaculatus* as the prey, and the predator mite, *Typhlodromus occidentalis*, in which a number of oranges, arranged in trays in a two-dimensional manifold, were used as the medium on which the prey species could develop. In all his simple universes, the interaction did not persist for very long, either the predator or prey disappearing as in Gause's experiments with various species of Protozoa, to which we have referred. But, when in Huffaker's experiments partial barriers were included in his universe, the co-existence of the predator and prey was maintained for three successive oscillations of the system. This more complex environment, in which the partial barriers restricted the movements of the species from orange to orange arranged in a two-dimensional grid, made it less likely that the predator would be in contact with the prey at all positions simultaneously. In a stochastic model for this type of system it would be necessary to incorporate a spatial grid, with restricted or partially restricted movements of the individuals from cell to cell, as in the model designed by Bartlett (1957*b*) for studying theoretically the development of a measles epidemic in a community. In terms of the simplified models used here, the nearest approach one could make to such a system would be to consider that the fraction k of the prey population exposed to risk was not a constant, but a variable; although at first sight it is not easy to see what form of functional dependence of k on the relative numbers of the two species should be adopted in the model. Generally speaking, however, it seems likely from the results for constant k given in this section, that even in the case of variable k, provided that this variability is not too extreme, the likelihood of both species being able to co-exist would be greatly increased.

REFERENCES

BARTLETT, M. S. (1957*a*). On theoretical models for competitive and predatory biological systems. *Biometrika*, **44**, 27–42.

BARTLETT, M. S. (1957*b*). Measles periodicity and community size. *J.R. Statist. Soc.*, A, **120**, 48–70.

BARTLETT, M. S., GOWER, J. C. & LESLIE, P. H. (1960). A comparison of theoretical and empirical results for some stochastic population models. *Biometrika*, **47**, 1–11.

GAUSE, G. F. (1934). *The Struggle for Existence*. Baltimore: Williams and Wilkins.

GAUSE, G. F. (1935). Vérifications expérimentales de la théorie mathématique de la lutte pour la vie. *Actualités Sci.* no. 277, 1–61. Paris: Hermann.

HUFFAKER, C. B. (1958). Experimental studies on predation: dispersion factors and predator–prey oscillations. *Hilgardia*, **27**, 343–83.

LESLIE, P. H. (1948). Some further notes on the use of matrices in population mathematics. *Biometrika*, **35**, 213–45.

LESLIE, P. H. (1958). A stochastic model for studying the properties of certain biological systems by numerical methods. *Biometrika*, **45**, 16–31.

LESLIE, P. H. & GOWER, J. C. (1958). The properties of a stochastic model for two competing species. *Biometrika*, **45**, 316–30.

Appendix I

For convenience, the bivariate distributions of N_1 and N_2 in the region of the stationary state for the three systems, to which reference has been made in the text, are given all together in this appendix. These grouped frequency distributions were tabulated by hand from the typed lists of results for each series, the marginal totals being checked by the distributions of each species (grouped in pairs of integers, 0–, 2–, 4–, 6–, etc.) recorded by the machine. It should be noted, however, in this connexion that all the variances and covariances given in the text were calculated on the machine from the individual results in each series, and not from the grouped data. They are therefore, in a sense, exact. The following were the values of the parameters used in computing the processes.

Parameter	Table 1	Table 2	Table 3
λ_1	1·2574	1·25	1·25
b_1	0·2577	0·25	0·25
α_1	0·0005148	0·001	0·001
β_1	0·0018018	0·015	0·050
λ_2	1·1892	1·25	1·25
b_2	0·2521	0·25	0·25
α_2	0·2838	0·375	0·25
k	1·0	1·0	0·20

Table 1

N_2 \ N_1	80–	90–	100–	110–	120–	130–	140–	150–	160–	170–	180–	190–	200–	210–	220–	Total
56–	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	1
60–	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	0
64–	.	2	6	1	2	.	.	2	2	2	.	.	.	.	.	17
68–	.	1	.	3	1	3	2	3	.	1	.	1	.	.	.	15
72–	.	.	.	4	2	6	4	1	2	.	1	2	.	.	.	22
76–	.	.	1	2	7	5	7	5	4	1	.	2	.	.	.	34
80–	.	1	3	9	8	9	10	4	5	6	1	.	1	2	2	61
84–	.	2	.	8	3	11	8	8	9	6	1	2	.	3	.	61
88–	1	.	2	6	14	10	8	7	10	6	3	2	2	.	1	72
92–	.	.	3	8	7	12	10	13	14	4	5	3	4	1	.	84
96–	.	.	1	7	9	20	20	13	7	11	5	3	.	.	1	97
100–	.	1	2	2	5	13	12	15	13	6	5	1	2	2	.	79
104–	.	1	2	2	.	5	14	6	8	13	6	2	3	1	1	64
108–	.	.	2	4	5	8	9	13	4	2	7	2	.	.	.	56
112–	.	.	.	3	3	14	5	10	6	8	4	2	2	.	.	57
116–	.	.	.	.	2	2	8	5	5	3	3	.	1	.	.	29
120–	.	.	.	1	1	5	3	2	4	.	1	.	.	.	.	17
124–	.	.	.	1	2	2	1	3	5	4	.	.	.	.	.	18
128–	.	.	.	.	.	2	.	1	1	1	.	.	.	.	.	5
132–	.	.	.	.	1	1	1	1	3	.	.	.	.	.	.	7
136–	.	.	.	.	.	.	.	.	3	.	.	.	.	.	.	3
140–	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	1
Total	1	9	22	61	72	128	123	112	105	74	42	22	15	9	5	800

Table 2

N_2 \ N_1	0–	4–	8–	12–	16–	20–	24–	28–	32–	36–	40–	44–	48–	52–	56–	60–	64–	68–	Total
0–	.	1	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2
2–	.	3	2	4	2	1	.	.	.	.	.	.	.	.	.	.	.	.	12
4–	.	3	2	2	5	5	1	2	.	1	.	.	.	.	.	.	.	.	21
6–	.	.	4	4	8	3	6	2	.	.	.	1	.	.	1	.	.	.	29
8–	.	.	2	5	8	6	5	6	1	1	2	2	.	.	.	.	1	.	39
10–	.	1	1	8	16	15	11	12	1	1	1	1	1	.	.	.	.	.	69
12–	.	1	2	9	14	7	3	4	4	.	5	2	2	.	1	.	1	.	55
14–	.	.	2	4	9	8	5	3	5	.	6	.	1	1	.	2	.	.	46
16–	.	1	1	4	4	4	5	7	2	3	4	1	1	.	.	2	.	1	40
18–	.	.	.	2	.	4	4	3	5	4	2	1	.	1	.	.	.	.	26
20–	.	.	.	2	.	4	1	2	3	1	2	1	.	.	1	.	.	.	17
22–	.	.	.	.	2	2	.	1	3	2	1	.	.	1	.	.	.	.	12
24–	.	.	.	.	.	3	1	1	.	2	.	.	.	.	.	.	.	.	7
26–	.	.	.	.	.	1	1	3	.	.	.	.	2	.	.	.	.	.	7
28–	.	.	.	.	1	1	1	.	1	2	.	.	.	.	.	.	.	.	6
30–	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	0
32–	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	1
Total	0	10	17	44	69	64	45	46	25	17	23	9	7	3	3	4	2	1	389

Table 3

N_1 / N_2	60–	66–	72–	78–	84–	90–	96–	102–	108–	114–	120–	126–	132–	138–	144–	150–	156–	162–	168–	174–	Total
6–	.	.	.	.	.	.	.	.	.	.	.	.	1	1	.	.	.	.	.	.	2
8–	.	.	.	.	.	.	1	1	.	.	1	.	3	2	1	.	.	.	.	.	9
10–	.	1	.	2	.	3	2	3	3	4	3	1	.	.	1	2	.	2	.	.	27
12–	.	.	1	.	2	.	3	4	1	9	3	4	3	2	2	4	1	.	.	.	39
14–	.	3	1	.	1	4	5	5	6	5	7	6	9	8	4	.	.	.	.	.	64
16–	1	1	2	.	1	7	8	7	10	9	7	7	2	6	2	1	1	1	.	.	73
18–	1	1	.	.	1	4	9	11	10	11	8	8	11	11	4	5	2	.	.	.	97
20–	.	1	2	1	1	5	8	15	20	19	9	9	11	9	13	3	2	.	.	.	128
22–	1	.	.	1	3	4	9	16	14	17	13	11	17	16	14	3	.	2	.	.	141
24–	.	1	.	1	4	4	12	3	16	13	19	14	11	14	7	1	1	1	2	.	124
26–	.	.	.	.	.	2	9	11	10	16	10	15	15	6	7	7	1	.	.	1	110
28–	.	.	1	.	1	2	7	5	6	9	6	11	11	3	5	2	.	.	.	.	69
30–	.	.	.	.	.	2	2	5	1	4	2	9	1	7	4	.	.	.	.	.	37
32–	.	.	.	.	.	2	4	1	6	2	9	4	4	2	1	.	.	.	.	.	35
34–	.	.	.	.	.	1	.	1	4	2	7	3	2	2	.	.	.	.	1	.	23
36–	.	.	.	.	.	.	1	2	1	2	2	1	1	2	.	.	.	.	.	.	12
38–	.	.	.	.	.	.	.	.	.	1	2	1	1	.	2	.	.	.	.	.	7
40–	.	.	.	.	.	.	.	.	.	.	1	1	.	.	.	.	.	.	.	.	2
Total	3	8	7	5	14	40	80	90	108	123	109	105	103	91	67	28	8	6	3	1	999

Biometrika (1960), **47**, 3 *and* 4, *p.* 235
Printed in Great Britain

Hierarchical birth and death processes

I. Theory

BY GUNNAR BLOM
University of Lund, Sweden

SUMMARY

A hierarchical birth and death process is defined by dividing the states of a birth and death process into substates according to certain rules. Only processes with constant transition probabilities are considered. In § 1, some basic results concerning integer-valued Markov processes are reviewed. In § 2, a special hierarchical process is discussed, and in § 3 equilibrium probabilities of the states of this process are derived in terms of given transition probabilities. In § 4, brief attention is paid to another special case. In § 5, a more general hierarchical process is analysed.

The processes considered in §§ 2–5 are finite. In § 6, this restriction is removed, so that the process is allowed to adopt an infinite denumerable number of states. Sufficient conditions for this general process to enter into statistical equilibrium are given.

Some of the results presented in this paper may be obtained from the general theory of reversible Markov chains and processes. However, the paper will be self-contained to a large extent, and most of the results will be derived from first principles.

1. SOME BASIC RESULTS CONCERNING INTEGER-VALUED MARKOV PROCESSES

The hierarchical processes considered in this paper are special cases of a general integer-valued Markov process with constant transition probabilities. The number of states of the process is assumed to be finite or, more generally, denumerable. For the general theory of Markov processes, the reader is referred to Ch. VI of the book by Doob (1953), to several publications by Feller (e.g. 1940, 1950, 1957) and to Kendall & Reuter (1957), where references are also given to several other important papers.

Denote the states of the process by $E_0, E_1, \ldots$, and suppose that the transition $E_i \to E_j$ occurs between t and $t+\Delta t$ with the probability $a_{ij}\Delta t + o(\Delta t)$ $(i \neq j)$. Further, define

$$a_{ii} = -\sum_{\nu \neq i} a_{i\nu}.$$

Let $P_m(t)$ be the probability that at time t the process adopts the state E_m. Naturally, this probability depends on the initial state of the process at, say, $t = 0$. We shall say that the process *enters into equilibrium* when $t \to \infty$ if

$$\lim_{t \to \infty} P_m(t) = P_m \quad (m = 0, 1, \ldots),$$

where the limit does not depend on the initial state, and if, in addition,

$$\sum_{m=0}^{\infty} P_m = 1.$$

The quantities P_m will be called *equilibrium* or *steady state probabilities*.

The process is called *irreducible* if, with positive probability, any state can be reached from any other state. It has been pointed out by Jensen (1954, p. 40), that an irreducible process of the kind considered here enters into equilibrium if and only if the mean number of steps for recurrence of a state E_m is finite for any m, and, moreover, the mean recurrence time is also finite for any m.

This result is essentially a consequence of Feller's theory of recurrent events. The conditions can be given another and generally more useful formulation, which is due to Jensen.

Consider the infinite system of equations in the unknowns $x_0, x_1, \ldots$

$$\sum_{m=0}^{\infty} a_{mj} x_m = 0 \quad (j = 0, 1, \ldots). \tag{1}$$

THEOREM. *An irreducible Markov process with constant transition probabilities enters into equilibrium if and only if there exists a solution* $x_m > 0$ $(m = 0, 1, \ldots)$ *to the system* (1) *such that the two series* $\sum_{m=0}^{\infty} x_m$ *and* $\sum_{m=0}^{\infty} (-a_{mm}) x_m$ *are both convergent.*

In particular, if the number of states is finite, the system is of course also finite and comprises as many unknowns and equations as there are states. The process enters into equilibrium if and only if $x_m > 0$ for all m.

For a process which satisfies the conditions of the theorem it is, at least in principle, easy to determine the equilibrium probabilities. Indeed, P_m is proportional to x_m, which implies that the probabilities can be determined from the system

$$\left.\begin{aligned} \sum_{m=0}^{\infty} a_{mj} P_m &= 0 \quad (j = 0, 1, \ldots), \\ \sum_{m=0}^{\infty} P_m &= 1. \end{aligned}\right\} \tag{2}$$

2. A SPECIAL FINITE HIERARCHICAL PROCESS

We shall begin by considering an ordinary birth and death process with a finite number of states E_m $(m = 0, 1, \ldots, n)$. At any time-point t the process may change from one state to neighbouring states according to the following rules:

$$E_m \to E_{m+1} \text{ with probability } a_m \Delta t \quad (m = 0, \ldots, n-1),$$
$$E_m \to E_{m-1} \text{ with probability } b_m \Delta t \quad (m = 1, \ldots, n).$$

More exactly, this means that, if the system has adopted the state E_m at time t, it changes during $(t, t+\Delta t)$ with probability $a_m \Delta t + o(\Delta t)$ to E_{m+1}, etc.

We shall assume that a_m and b_m are positive for any m. The process is then irreducible and enters into equilibrium when $t \to \infty$. The equilibrium probabilities P_m are obtained from the general system of equations (2), which takes the form

$$\begin{aligned} a_0 P_0 &= b_1 P_1, \\ (a_m + b_m) P_m &= a_{m-1} P_{m-1} + b_{m+1} P_{m+1} \quad (m = 1, \ldots, n-1), \\ b_n P_n &= a_{n-1} P_{n-1}. \end{aligned}$$

The system has the solution

$$P_m = \frac{a_0 a_1 \ldots a_{m-1}}{b_1 b_2 \ldots b_m} P_0, \tag{3}$$

where P_0 is determined by observing that the probabilities add up to unity.

So far everything is well known. To arrive at the special hierarchical process considered in this section, we shall (*a*) divide the states of the birth and death process into certain substates, (*b*) specify what will be called the permissible states. We shall give a detailed description of these two parts of the definition in the remaining part of the section.

It should be pointed out that the process considered here has mainly theoretical interest, but it is of basic importance for the treatment of the more general process introduced in § 5.

(*a*) *Division of the birth and death process into substates*

E_0 is left unchanged. E_1 is divided into N substates E_i such that

$$E_0 \to E_i \text{ with probability } a_0 p_i \Delta t \quad (i = 1, \ldots, N),$$
$$E_i \to E_0 \text{ with probability } b_1 q_i \Delta t \quad (i = 1, \ldots, N).$$

To make the system of notations wholly consistent, it would perhaps be better to write E instead of E_0, but we prefer to use the usual notation for the null state.

Furthermore, E_2 is divided into N^2 substates E_{ij} such that

$$E_i \to E_{ij} \text{ with probability } a_1 p_j \Delta t,$$
$$E_{ij} \to E_i \text{ with probability } b_2 q_j \Delta t.$$

It should be observed that the 'birth' probability does not depend on i, only on j.

Generally, E_m is divided into N^m substates $E_{ij \ldots yz}$ (m indices) such that

$$E_{ij \ldots y} \to E_{ij \ldots yz} \text{ with probability } a_{m-1} p_z \Delta t,$$
$$E_{ij \ldots yz} \to E_{ij \ldots y} \text{ with probability } b_m q_z \Delta t.$$

With respect to the quantities p_ν and q_ν we shall make the following assumption.

ASSUMPTION 1. *The quantities $p_1, \ldots, p_N$ and $q_1, \ldots, q_N$ are positive.*

The number of indices in $E_{ij \ldots yz}$ will be termed the *order* of the state.

Note that the total number of substates resulting from the successive divisions is $1 + N + \ldots + N^n$. However, some of these substates may sometimes be excluded, as will appear from the second part of the definition.

(*b*) *Specification of permissible states*

When the transition probabilities mentioned above are given, the hierarchical process would be uniquely defined, were it not for the fact that we shall sometimes restrict the process so as to adopt only certain states, which will be called *permissible*. This means that, in the above definition of a transition from a state to a next higher one, we shall allow z to range only over a subsequence of the integers $1, \ldots, N$. The subsequence may depend on the previous state $E_{ij \ldots y}$. Using another terminology, we may say that the p_ν are sometimes put equal to zero. The states which the process cannot adopt will be termed *forbidden*.

The process is defined when a complete list of permissible states is given. Such a list may take on different shapes depending on the problem under study. However, in order to arrive at simple and useful results, we shall make the following rather general assumptions.

ASSUMPTION 2. *There is at least one permissible state of each order.*

ASSUMPTION 3. *If a state $E_{ij \ldots z}$ is permissible, then also the states obtained by permuting the indices in all possible ways are permissible.*

ASSUMPTION 4. *If a state $E_{ij\dots z}$ is permissible, then also the states obtained by deleting any number of indices are permissible.*

Strictly speaking, Assumption 3 is not needed for the analysis of the special process considered in this section, but we shall need it later.

In Part II of the paper, we shall encounter an important example of a situation where only certain states are permissible. In fact, we shall then require that those states are permissible for which all the indices are distinct. Clearly, the number of permissible states is then reduced from the maximum number

$$1+N+N^2+\dots+N^n \quad \text{to} \quad 1+N+N(N-1)+\dots+N!.$$

When the four sequences $\{a_\nu\}$, $\{b_\nu\}$, $\{p_\nu\}$, $\{q_\nu\}$ and the permissible states are specified, the hierarchical birth and death process is completely defined. We may somewhat loosely describe it as follows. If the system starts from the null state E_0 and passes successively to higher states, it 'remembers' all the substates where it has been before. If 'death' occurs, the system passes to a next lower state which is uniquely defined by the previous state. This remark about the 'memory' of the process should not give the reader the false impression that the process is non-Markovian. In fact, if the substates are renumbered using a single sequence of indices, it is recognized that any transition depends only on the last state and not on any previous states. Thus the process is Markovian.

3. STEADY STATE PROBABILITIES OF THE SPECIAL PROCESS

Consider the special hierarchical birth and death process defined in the preceding section. Our main task in this section will be to derive expressions for the steady state probabilities of various states $E_{ij\dots yz}$. For this purpose, we could use the system (2) directly. However, it is more informative to start from first principles.

Suppose that the process begins in E_0 at $t=0$, and let $P_{ij\dots z}(t)$ denote the probability that at time t it assumes the state $E_{ij\dots z}$. Let the number of indices in these expressions be m. By comparing two adjacent time-points t and $t+\Delta t$, we obtain the relations

$$P_{ij\dots yz}(t+\Delta t) = (1-a_m\Sigma p_\nu - b_m q_z)\,P_{ij\dots yz}(t) + a_{m-1}p_z P_{ij\dots y}(t) + b_{m+1}\Sigma q_\nu P_{ij\dots zv}(t).$$

Denoting by $P_0(t)$ the probability that the system is in E_0 at time t, we also have the relation

$$P_0(t+\Delta t) = (1-a_0\Sigma p_\nu)\,P_0(t) + b_1\Sigma q_\nu P_\nu(t).$$

In these relations Σ denotes summation over all ν such that $E_{ij\dots zv}$ is a permissible state. The first relation holds for any number of indices $ij\dots yz$ from 1 to $n-1$. It also holds when this number is equal to n, provided that we put $a_n = b_{n+1} = 0$.

Note the importance of the assumptions made in § 2 about the choice of permissible states. By Assumption 2, there is at least one relation corresponding to any order m. Furthermore, by Assumption 4, the transitions $E_{ij\dots yz} \to E_{ij\dots y}$ and $E_{ij\dots zv} \to E_{ij\dots z}$ are always possible (if this were not so, one or perhaps several terms would disappear and the equation would not hold).

We now apply the general theory reviewed in § 1. It follows from Assumptions 1 and 2 that the process is irreducible and enters into equilibrium when $t \to \infty$. Letting $\Delta t \to 0$ in the above relations and putting the resulting derivatives equal to zero, we obtain the steady state system of equations

$$a_0\Sigma p_\nu P_0 = b_1\Sigma q_\nu P_\nu,$$

$$a_m\Sigma p_\nu P_{ij\dots yz} + b_m q_z P_{ij\dots yz} = a_{m-1}p_z P_{ij\dots y} + b_{m+1}\Sigma q_\nu P_{ij\dots zv}. \tag{4}$$

The solution of this seemingly complicated system turns out to be very simple. As is readily verified by substitution, the solution is obtained by equating the second term in the left member of (4) to the first term in the right member, which yields

$$P_{ij\dots yz} = \frac{a_{m-1}}{b_m}\frac{p_z}{q_z}P_{ij\dots y}.$$

This recursive formula leads to

$$P_{ij\dots yz} = \frac{a_0 a_1 \dots a_{m-1}}{b_1 b_2 \dots b_m}\frac{p_i p_j \dots p_z}{q_i q_j \dots q_z}P_0. \tag{5}$$

We conclude from the form of this expression that the steady state probability is the same for all permutations of the letters $ij \dots yz$. The factor P_0 is as yet undetermined. Since the probabilities add up to unity, we have

$$\frac{1}{P_0} = 1 + \Sigma\frac{a_0 \dots a_{m-1}}{b_1 \dots b_m}\frac{p_i \dots p_z}{q_i \dots q_z}, \tag{6}$$

with summation over all sequences of $1, 2, \dots, n$ indices $ij \dots yz$ corresponding to permissible states $E_{ij\dots yz}$. This compact formula may be difficult to interpret, and the reader may prefer the alternative formula given below.

We shall sometimes be interested in considering the totality of all permissible states $E_{ij\dots z}$ of a given order m. We shall denote this 'super-state' by $\mathscr{E}_m$ $(m = 0, 1, \dots, n)$. Naturally $\mathscr{E}_0 = E_0$. The steady state probability corresponding to $\mathscr{E}_m$ will be designated by Π_m. To find Π_m we have of course only to sum (5) over all permissible states of order m. With a slight change of notation we have

$$\Pi_m = \frac{a_0 \dots a_{m-1}}{b_1 \dots b_m}\quad\frac{p_{\nu_1} \dots p_{\nu_m}}{q_{\nu_1} \dots q_{\nu_m}}P_0 \tag{7}$$

with summation over as many sequences of indices $\nu_1, \dots, \nu_m$ as there are permissible states of order m. We may then write equation (6) in the alternative form

$$\frac{1}{P_0} = 1 + \sum_{m=0}^{n}\Pi_m.$$

We shall end the section by considering a very special case. Let us assume that *all* states are permissible and, moreover, that

$$\sum_{\nu=1}^{N}\frac{p_\nu}{q_\nu} = 1.$$

Then the sum in the right member of (7) becomes $(\Sigma p_\nu/q_\nu)^m = 1$, and hence this formula reduces to

$$\Pi_m = \frac{a_0 \dots a_{m-1}}{b_1 \dots b_m}P_0.$$

Consequently, as seen from (3), the probabilities Π_m are in this special case identical with the steady state probabilities P_m of the simple birth and death process which served as our starting-point in § 2 when defining the hierarchical process.

4. Another special hierarchical process

As pointed out earlier, the process considered in the preceding two sections has mainly theoretical interest. From a practical point of view a more interesting case arises if the transition rules given in § 2 are changed in the following way:

$E_{ij\dots y} \to E_{ij\dots yz}$ with probability $a_{m-1}p_z\Delta t$,
$E_{ij\dots yz} \to E_{j\dots yz}$ with probability $b_m q_i \Delta t$.

Thus, when passing to a lower state, the first index is deleted instead of the last one. In all other respects the rules are the same as before. The resulting hierarchical process has interesting applications to queueing theory (cf. Part II).

By the method used in § 3 it can be proved that the steady state probabilities of the new process are given by the same expressions (5) as before, provided that

$$q_1 = q_2 = \ldots = q_N.$$

However, if the q_ν are unequal, no such simple solution is generally possible. To prove this statement, we proceed as follows.

Suppose that Assumptions 1–4 hold. By analogy to (4) we have the system of equations

$$a_m \Sigma p_\nu P_{ij\ldots yz} + b_m q_i P_{ij\ldots yz} = a_{m-1} p_z P_{ij\ldots y} + b_{m+1} \Sigma q_\nu P_{\nu ij\ldots z}.$$

It is easy to verify that if the probabilities P satisfy (5), then the first term in the left member and the second term in the right member cancel out; hence the system reduces to

$$b_m q_i P_{ij\ldots yz} = a_{m-1} p_z P_{ij\ldots y}.$$

Inserting the expressions (5) in this relation, we obtain after reduction $q_i = q_z$. Consequently, if the q_ν are equal, the system is certainly satisfied by these simple expressions. It also follows that if in the list of permissible states $E_{ij\ldots z}$ two of the indices, say i and z, may assume any of the $N(N-1)$ pairs of distinct integers $(1,2)$, $(1,3)$, …, $(N-1, N)$, then (5) holds *if and only if* all the q_ν are equal.

It may be added that, since (5) holds in this case, this is of course also true of the expression (7) for the super-state probability Π_m.

5. THE GENERAL FINITE HIERARCHICAL PROCESS

We shall now introduce a general process, which contains the two processes discussed in the preceding sections as special cases. The division of the original birth and death process into substates is made exactly as before; only the transition rules are affected by the generalization.

The transitions of the general process may be schematically described as follows (as before, m is the number of indices in $E_{ij\ldots yz}$ and ranges from 0 to n; the indices themselves range from 1 to N):

$$\begin{aligned} E_{ij\ldots y} \to & \begin{cases} E_{zij\ldots y}, & a_{m-1,1}\, p_z \Delta t, \\ E_{izj\ldots y}, & a_{m-2,2}\, p_z \Delta t, \\ \ldots\ldots\ldots & \ldots\ldots\ldots \\ E_{ij\ldots yz}, & a_{m-1,m}\, p_z \Delta t; \end{cases} \\ E_{ijk\ldots yz} \to & \begin{cases} E_{jk\ldots yz}, & b_{m,1}\, q_i \Delta t, \\ E_{ik\ldots yz}, & b_{m,2}\, q_j \Delta t, \\ \ldots\ldots\ldots & \ldots\ldots\ldots \\ E_{ij\ldots y}, & b_{m,m}\, q_z \Delta t. \end{cases} \end{aligned} \tag{8}$$

(The quantities in the right-hand column denote the probabilities corresponding to the various transitions.) In short, transitions may occur between states which have all but one index in common. Expressed otherwise, a next higher state is obtained by adding one index in any position, and a next lower state by deleting one index in any position.

To avoid ambiguity we shall demonstrate by an example how the rules should be interpreted. Consider transitions from states of order $m = 2$ to the next lower states. For example, the transition $E_{21} \to E_1$ takes place with probability $b_{21} q_2 \Delta t$ in the interval $(t, t+\Delta t)$, and $E_{12} \to E_1$ with probability $b_{22} q_2 \Delta t$. However, since E_{11} turns into E_1 by deleting either the first or the second index, the transition $E_{11} \to E_1$ occurs with probability $(b_{21} + b_{22}) q_1 \Delta t$.

We shall always suppose that the general process satisfies Assumptions 1–4 of § 2 and, in addition, the following assumption:

ASSUMPTION 5. *The quantities $a_{m-1,\nu}$ and $b_{m\nu}$ $(\nu = 1, \ldots, m;\ m = 1, \ldots, n)$ are non-negative. For any fixed m, at least one $a_{m-1,\nu}$ and at least one $b_{m\nu}$ is positive.*

A process defined by the transition rules (8) and satisfying Assumptions 1–5 will be called a *finite hierarchical birth and death process*. Any such process is irreducible and enters into equilibrium (cf. § 1).

It will be convenient to introduce the notations

$$a_{m-1} = \sum_{\nu=1}^{m} a_{m-1,\nu}; \quad b_m = \sum_{\nu=1}^{m} b_{m\nu} \quad (m = 1, \ldots, n). \tag{9}$$

By Assumption 5, a_{m-1} and b_m are positive for any m.

The equilibrium probabilities of the states $E_{ij\ldots yz}$ are determined by the system of linear equations

$$\begin{aligned} a_m \Sigma p_\nu P_{ij\ldots yz} + (b_{m1} q_i + \ldots + b_{mm} q_z) P_{ij\ldots yz} \\ = [a_{m-1,1} p_i P_{j\ldots yz} + \ldots + a_{m-1,m} p_z P_{ij\ldots y}] \\ + [b_{m+1,1} \Sigma q_\nu P_{\nu ij\ldots z} + \ldots + b_{m+1,m+1} \Sigma q_\nu P_{ij\ldots z\nu}]. \end{aligned} \tag{10}$$

In each sum the summation should be extended over all permissible states. By Assumptions 3 and 4, each sum ranges over the same subsequence of the integers $1, \ldots, N$. For $m = 0$ and $m = n$, the system should be modified in the usual way; for brevity we do not reproduce these equations.

In the general case the solution of the system (10) is complicated and will not be reproduced here. However, it sometimes turns out to be quite simple. In particular, we are interested in finding out under what conditions the solution is furnished by the simple expressions (5) encountered in the previous sections. For this purpose, we simply substitute these expressions in the system. It is seen that the first term in the left member and the sum within the second pair of brackets then cancel out, and after some reduction the following condition for the validity of (5) is obtained

$$\frac{a_{m-1,1} q_i + \ldots + a_{m-1,m} q_z}{b_{m,1} q_i + \ldots + b_{m,m} q_z} = \frac{a_{m-1}}{b_m}. \tag{11}$$

This condition is satisfied in two main cases. The corresponding processes will be termed hierarchical processes of Type A and B, as described below.

DEFINITION 1. *If Assumptions 1–5 hold and if, in addition,*

$$q_1 = q_2 = \ldots = q_N, \tag{12}$$

then the hierarchical process is said to be of Type A.

The condition (11) is evidently fulfilled in this case.

DEFINITION 2. *If Assumptions* 1–5 *hold and if, in addition, the quantities* $a_{m-1,\nu}$ *and* $b_{m\nu}$ *are proportional for any fixed* m, i.e.

$$a_{m-1,\nu} = c_m b_{m\nu} \quad (c_m > 0;\ \nu = 1, \ldots, m), \tag{13}$$

then the hierarchical process is said to be of Type B.

(Note that (13) implies that if $a_{m-1\nu}$ is zero, then also $b_{m\nu}$ is zero, and vice versa.) Also in this case condition (11) is satisfied.

We infer from these considerations that the steady state probabilities of hierarchical processes of Type A and B are given by the expressions (5). As a consequence, also the expressions (7) for the steady state probabilities of the super-states Π_m are valid for these processes. These results greatly facilitate the study of Type A and Type B processes.

In § 4 we have already seen an example of a hierarchical process of Type A. The special process discussed in §§ 2 and 3 is obtained by taking $a_{m-1\,m} = a_{m-1}$, $b_{mm} = b_m$ and equating the other $a_{m-1\nu}$ and $b_{m\nu}$ to zero. Hence this process is of Type B. Further examples will be presented in Part II of the paper.

It might be added that certain of the results proved in this section are closely connected with the theory of reversible Markov chains and processes put forward by Kolmogorov (1936), Kendall (1958) and Kramer (1959). A reader familiar with this theory will find that a hierarchical process of Type B (but not Type A) is reversible.

6. THE GENERAL INFINITE HIERARCHICAL PROCESS

The processes introduced in the foregoing section have a finite number of states. We shall now remove this restriction. It will be supposed that the number n of states of the original birth and death process may be infinite. As before, the states E_m ($m = 0, 1, \ldots$) are divided into a *finite* number N^m of substates. Also in other respects the definitions are the same as before, i.e. the transitions take place in accordance with the scheme (8), and Assumptions 1–5 are supposed to hold. The resulting process will be called an *infinite hierarchical birth and death process*. The classification into Type A and Type B processes is extended also to the infinite case (cf. § 5).

As before, there are two problems of interest. First, does an infinite process enter into equilibrium when $t \to \infty$? Secondly, what are the equilibrium probabilities?

The second question is easily answered. Provided that the first question is answered in the positive, it is realized that the equilibrium probabilities satisfy the system (10), which is now infinite. In particular, if the conditions (12) or (13) are fulfilled, the solution (5) is obtained. It remains to discuss the first question.

The behaviour of the process when $t \to \infty$ is determined by the properties of the two double infinite sequences $\{a_{m-1,\nu}\}$ and $\{b_{m\nu}\}$ ($\nu = 1, \ldots, m$; $m = 1, 2, \ldots$, ad inf.). A sufficient set of conditions for the process to enter into equilibrium may be obtained from the theorem of § 1 in the following way.

Consider the infinite system of equations (cf. the system (10))

$$\begin{aligned} a_m \Sigma p_\nu u_{ij\ldots yz} + (b_{m1} q_i + \ldots + b_{mm} q_z)\, u_{ij\ldots yz} \\ = [a_{m-11} p_i u_{j\ldots yz} + \ldots + a_{m-1\,m} p_z u_{ij\ldots y}] \\ + [b_{m+11} \Sigma q_\nu u_{\nu ij\ldots z} + \ldots + b_{m+1\,m+1} \Sigma q_\nu u_{ij\ldots z\nu}]. \end{aligned} \tag{14}$$

Note that for $m = 0$ the system is simplified to

$$a_0 \Sigma p_\nu u_0 = b_1 \Sigma q_\nu u_\nu.$$

In each sum, the letter ν ranges over the subsequence of $1, \ldots, N$ corresponding to the permissible states.

THEOREM 1. *Consider an infinite hierarchical birth and death process which satisfies Assumptions* 1–5. *The process enters into equilibrium when* $t \to \infty$ *if there exists a solution* $u_{ij \ldots yz} > 0$ *of the system* (14) *such that*

$$\sum_{m=0}^{\infty} U_m < \infty, \tag{15}$$

$$\sum_{m=0}^{\infty} (a_m + b_m) U_m < \infty, \tag{16}$$

where U_m *is the sum of all* $u_{ij \ldots yz}$ *corresponding to permissible states of order* m, *and* a_m *and* b_m *are defined by* (9).

Proof. It follows from the theorem of § 1 that the process enters into equilibrium if and only if

$$\Sigma u_{ij \ldots yz} < \infty,$$

$$\Sigma(a_m \Sigma p_\nu + b_{m1} q_i + \ldots + b_{mm} q_z) u_{ij \ldots yz} < \infty$$

with summation over all permissible states. The first condition is identical with (15). With respect to the second series, we observe that since the number of quantities p_ν and q_ν is finite, a quantity C can be found such that the factor within the parentheses is less than $C(a_m + b_m)$. Consequently, the second series certainly converges if (16) holds. This proves the theorem.

In the particular case when the process is of Type A or B, a more convenient sufficient set of conditions is given by

THEOREM 2. *Consider an infinite hierarchical birth and death process of Type A or B. Let* a_m *and* b_m *be defined by* (9) *and set*

$$\sum_{\nu=1}^{N} \frac{p_\nu}{q_\nu} = c.$$

If

$$\sum_{m=1}^{\infty} \frac{a_0 \ldots a_{m-1}}{b_1 \ldots b_m} c^m < \infty \tag{17}$$

and

$$\sum_{m=1}^{\infty} (a_m + b_m) \frac{a_0 \ldots a_{m-1}}{b_1 \ldots b_m} c^m < \infty, \tag{18}$$

then the process enters into equilibrium when $t \to \infty$.

Proof. Condition (15) becomes in this case

$$\Sigma \frac{a_0 \ldots a_{m-1}}{b_1 \ldots b_m} \frac{p_i \ldots p_z}{q_i \ldots q_z} < \infty$$

with summation over all permissible states. This condition certainly holds if the series obtained by summing the same expression over all states (permissible and forbidden) is convergent, i.e. if (17) holds. Similarly, it follows that condition (16) holds if condition (18) is fulfilled. This proves the theorem.

REFERENCES

DOOB, J. L. (1953). *Stochastic Processes.* New York: John Wiley and Sons.

FELLER, W. (1940). On the integro-differential equations of purely discontinuous Markoff processes. *Trans. Amer. Math. Soc.* **48**, 488.

FELLER, W. (1950). *An Introduction to Probability Theory and its Applications*, **1**. New York: John Wiley and Sons.

FELLER, W. (1957). On boundaries and lateral conditions for the Kolmogorov differential equations. *Ann. Math.* **65**, 527.

JENSEN, A. (1954). *A Distribution Model.* Copenhagen.

KENDALL, D. G. (1958). Integral representations for Markov transition probabilities. *Bull. Amer. Math. Soc.* **64**, 358.

KENDALL, D. G. & REUTER, G. E. H. (1957). The calculation of the ergodic projection for Markov chains and processes with a countably infinity of states. *Acta Math.* **97**, 103.

KOLMOGOROV, A. (1936). Zur Theorie der Markoffsken Ketten. *Math. Ann.* **112**, 155.

KRAMER, H. P. (1959). Symmetrizable Markov Matrices. *Ann. Math. Statist.* **30**, 149.

Biometrika (1960), **47**, 3 *and* 4, *p.* 245
Printed in Great Britain

Hierarchical birth and death processes

II. Applications

BY GUNNAR BLOM
University of Lund, Sweden

SUMMARY

In Part I of this paper (pp. 235–44), the theory of hierarchical birth and death processes was expounded. In the present part, two kinds of applications will be given, namely, to machine interference and telephone trunking problems.

In § 1, a review is given of the well-known queueing problem arising when a group of identical machines is serviced by one operator. Stoppages are assumed to follow the simple Poisson scheme, and the service-time distribution is supposed to be exponential. In § 2, an analysis is made of the corresponding problem for machines with different needs of service but the same service-time distribution. This problem does not seem to have been discussed before in the literature. In § 3, the result is generalized to the case of several operators.

Palm (1937) determined the equilibrium probabilities of the states of a busy-signal telephone system used by a finite number of subscribers with different call frequencies. It is shown in § 4 that this problem can be solved by means of the theory of hierarchical birth and death processes.

1. SERVICING OF MACHINES WITH THE SAME NEED OF SERVICE

For later purposes we shall sum up some properties of a well-known queueing process considered by Palm (1947) and later by Feller (1950, chap. 17.7), and several other writers.

Let n machines be serviced by one operator and suppose that the machines behave in the same way as follows. If a machine is working at time t, the probability is $\lambda\Delta t+o(\Delta t)$ that it stops before $t+\Delta t$. Also, if it is being serviced at time t, the probability is $\mu\Delta t+o(\Delta t)$ that service ends before $t+\Delta t$. The machines are attended to in the order they stop.

The resulting queueing process is of the simple birth and death type mentioned at the beginning of § 2 of Part I. Denote by P_m the steady state probability of the event 'm machines are being serviced or are waiting for service'. To determine P_m we may use formula (3) in the section just mentioned with

$$a_m = (n-m)\lambda; \quad b_m = \mu.$$

This leads to

$$P_m = n(n-1)\dots(n-m+1)\rho^m P_0 \quad (m = 1, \dots, n), \tag{1}$$

where

$$\rho = \frac{\lambda}{\mu}$$

is the servicing factor. Moreover

$$\frac{1}{P_0} = 1 + \sum_{m=1}^{n} n(n-1)\dots(n-m+1)\rho^m.$$

Clearly, P_0 is the fraction of time during which the operator is free. We shall also be interested

in determining the fractions of time S, M and W during which a certain machine is being serviced, is in working state, or is waiting for service, respectively. We have

$$\left.\begin{aligned} S &= \frac{1}{n}\sum_{m=1}^{n} P_m = \frac{1-P_0}{n}, \\ M &= \frac{1}{n}\sum_{m=0}^{n}(n-m)P_m, \\ W &= 1-S-M. \end{aligned}\right\} \tag{2}$$

For numerical purposes it is convenient to use the following formula for the machine efficiency M, viz.

$$M = \frac{1-P_0}{n\rho}.$$

However, in the sequel we shall be more interested in a somewhat different expression (cf. Benson & Cox (1951), p. 70). Let us write $P_0(n)$ instead of P_0 in order to emphasize that it signifies the event 'operator in charge of n machines is free'. Since by (1)

$$\begin{aligned} \sum_{m=0}^{n}(n-m)P_m &= n\left[1+\sum_{m=1}^{n-1}(n-1)(n-2)\dots(n-m)\rho^m\right]P_0(n) \\ &= n\frac{P_0(n)}{P_0(n-1)}, \end{aligned}$$

we obtain from (2)

$$M = \frac{P_0(n)}{P_0(n-1)}. \tag{3}$$

Thus the machine efficiency can be expressed as the quotient between $P_0(n)$ and the probability $P_0(n-1)$ that an operator attending $n-1$ machines is free.

It is instructive to examine a little more closely what this means. For this purpose, set $P_0(n) = MP^*$. By the definition of conditional probability, P^* is the conditional probability that, when it is known that one machine is in working state, also all the others are working. The relation just proved states that $P^* = P_0(n-1)$; hence intuition does not lead us astray if, during the fraction of time when a certain machine is running, we disregard this machine and consider the remaining $n-1$ machines as a closed system. It should be observed that when the process is non-Markovian, this reasoning is generally not correct.

2. Servicing of machines with different needs of service. One operator

It sometimes happens that, of a group of machines, some require more attention than others. For example, in the textile industry it is, at least in Europe, common practice that different types of cloth are woven on the looms attended by the same weaver. Some machines may then stop more frequently than others, for example, because the end-breakage rates depend on the count of the yarn, the design of the fabric, etc., and hence tend to be different in the long run for different cloths. Also the average time required for attending to a loom when a stop occurs may depend upon the cloth woven; however, this dependence is generally insignificant, and we shall disregard it altogether in this section; i.e. we shall suppose that the service-time distribution is the same for all the machines in the group.

We make then the following assumptions. The group consists of n machines. The probability is $\lambda_\nu \Delta t + o(\Delta t)$ that the νth machine calls for service between t and $t+\Delta t$. Also, if a machine is being attended to, the probability is $\mu \Delta t + o(\Delta t)$ that service ends between

t and $t+\Delta t$. We wish to calculate the fraction of time P_0 when the operator is free, and also the fractions of time S_ν, M_ν and W_ν when the νth machine is being attended to, is in working state, and is waiting for service, respectively.*

The solution to the problem is readily obtained from the general theory of Part I. Denote by $E_{ij\dots z}$ the event 'the ith machine is being attended to, and the jth, ..., zth machines are queueing for service in this order', and by $\mathscr{E}_m$ the event 'm machines are not working'. The corresponding equilibrium probabilities are denoted by $P_{ij\dots z}$ and Π_m, respectively. Clearly, the indices $ij\dots z$ are always distinct, and hence the total number of permissible states equals $1+n+n(n-1)+\dots+n!$. The following transitions are possible:

$$E_{ij\dots y} \to E_{ij\dots yz} \text{ with probability } \lambda_z \Delta t,$$

$$E_{ij\dots yz} \to E_{j\dots yz} \text{ with probability } \mu \Delta t.$$

Further, Assumptions 1–4 (cf. § 2 of Part I) are fulfilled, and we note that the process is of the special type described in § 4 of Part I. In fact, in the scheme given in that section we may take $N = n$ and

$$a_{m-1} = b_m = 1 \quad (m = 1, \dots, n); \qquad p_\nu = \lambda_\nu, \quad q_\nu = \mu \quad (\nu = 1, \dots, n).$$

By the terminology introduced in § 5 of Part I, the queueing process is a hierarchical birth and death process of Type A.

By the basic formula (5) in Part I, the steady state probabilities are given by

$$P_{ij\dots z} = \rho_i \rho_j \dots \rho_z P_0, \tag{4}$$

where

$$\rho_\nu = \frac{\lambda_\nu}{\mu}$$

is the servicing factor of the νth machine.

Next we shall write down an expression for the probability Π_m that m machines are not working. By formula (7) in Part I

$$\Pi_m = \Sigma \rho_i \rho_j \dots \rho_z P_0 = m! \, \Sigma \rho_1 \rho_2 \dots \rho_m P_0,$$

where the sign Σ in the latter expression denotes summation of the $\binom{n}{m}$ terms obtained by choosing m indices at a time from the sequence $1, \dots, n$. Furthermore, by adding the probabilities Π_m we find

$$\frac{1}{P_0} = 1 + \Sigma \rho_1 + 2! \, \Sigma \rho_1 \rho_2 + \dots + n! \, \rho_1 \dots \rho_n. \tag{5}$$

The fraction of time S_ν when the νth machine is attended to is obtained by adding all $P_{ij\dots z}$ with first index $i = \nu$. Hence by (4)

$$S_\nu = \rho_\nu (1 + \Sigma' \rho_1 + 2! \, \Sigma' \rho_1 \rho_2 + \dots) P_0,$$

where Σ' denotes summation over all indices $1, \dots, n$ except ν.

Similarly, the fraction of time M_ν when the νth machine is doing productive work, i.e. its machine efficiency, is obtained by adding over all sequences $ij\dots z$ which do not contain ν.

* As in § 1 we suppose that the machines are attended in the order in which they stop. This assumption is now rather special; usually it is more economical to give priority to machines with small λ_ν.

Hence
$$M_\nu = (1+\Sigma'\rho_1+2!\,\Sigma'\rho_1\rho_2+\ldots)\,P_0, \tag{6}$$
where Σ' has the same meaning as before. Comparing the last two expressions, we find
$$S_\nu = \rho_\nu M_\nu. \tag{7}$$
Moreover, using (5) we obtain the following generalization of (3)
$$M_\nu = \frac{P_0(1,\ldots,n)}{P_0(1,\ldots,\nu-1,\nu+1,\ldots,n)}. \tag{8}$$
This relation reveals an interesting fact, viz. that the comments made in the last paragraph of § 1 are valid even in this more general situation.

It is now clear how calculations should be made in practice. The quantities ρ_ν may be supposed to be known in advance from time study, analysis of production figures, or the like. From (5) we calculate P_0 for the whole group of n machines and also for the n groups of $n-1$ machines obtained by deleting one value ν at a time. Thus $n+1$ such values are obtained in all. Next, the machine efficiencies are calculated by means of (8) for each machine, and then S_ν is obtained from (7). As a check on the computations, one may use the relation
$$\Sigma S_\nu = 1-P_0(1,\ldots,n),$$
which expresses the obvious fact that the sum of the fractions of time when the machines are attended to equals the probability that at least one machine is not working. Finally, since $S_\nu+M_\nu+W_\nu = 1$, the waiting-times W_ν can be determined, and the problem is completely solved.

There are some general conclusions which may be drawn from these results. First, it follows from (5) that, if the ρ_ν are given all possible values such that $\Sigma\rho_\nu$ is constant, the fraction of time P_0 when the operator is free becomes smallest when the ρ_ν are equal. Consequently, from the operator's point of view, the latter situation is the most unfavourable one.

Secondly, if the ρ_ν vary such that P_0 is held constant, the average machine efficiency $\bar{M}$ is smallest when the ρ_ν are equal. For by (6) we have after some reduction
$$\bar{M} = P_0\left[1+\left(1-\frac{1}{n}\right)\Sigma\rho_1+\left(1-\frac{2}{n}\right)2!\,\Sigma\rho_1\rho_2+\ldots\right].$$
It can be proved that when P_0 and hence the expression in the right member of (5) is held constant, $\bar{M}$ has a minimum for constant ρ_ν. This result has perhaps some practical relevance in the case where there are several possibilities for assigning machines to a worker, and it is required to find a combination which, given a suitable work load $1-P_0$, makes total production as large as possible. Evidently, our result implies that machines with the same or nearly the same ρ_ν should not be combined into a group. This remark gives rise to many interesting questions, which are outside the scope of this paper.

3. Servicing of machines with different needs of service. Several operators

The solution of the one-operator problem considered in the previous section will now be generalized to the case of several operators. Suppose that the group of n machines is serviced by $r > 1$ operators. In all other respects the conditions are the same as before, i.e. the probability is $\lambda_\nu\Delta t+o(\Delta t)$ that the νth machine stops between t and $t+\Delta t$, and the

probability is $\mu\Delta t + o(\Delta t)$ that a machine which is being serviced returns to working state between t and $t+\Delta t$.

Denote by $E_{ij\ldots z}$ the state 'the ith, jth, ..., zth machines have stopped in this order and are being serviced or are waiting for service'. (Note that this statement does not necessarily imply that the machines stopped in immediate succession; it may well happen that one or several machines stopped, for example, after the ith and before the jth one and that service of these machines has already finished.)

It is easy to verify that the transitions from one state to the other obey the rules of the general hierarchical process considered in § 5 of Part I. First, a transition to a next higher state occurs when a machine stops; evidently we have $E_{ij\ldots y} \to E_{ij\ldots yz}$ with probability $\lambda_z \Delta t$. Accordingly, in the first part of the scheme (8) given in that section we take

$$p_\nu = \lambda_\nu; \quad a_{m-1\nu} = \begin{cases} 0 & \text{for} \quad 1 \leqslant \nu \leqslant m-1, \\ 1 & \text{for} \quad \nu = m. \end{cases}$$

Secondly, when the process passes to a lower state, there are two possibilities: If the number of indices m in $E_{ij\ldots z}$ is less than or equal to r, all the inactive machines are cared for at the same time; hence a transition is possible to any of the m lower states obtained by deleting one index at a time. On the other hand, if m is greater than r, only the r machines which stopped first are being attended to; consequently, only the indices in the first r positions may then be deleted when passing to a next lower state. Accordingly, in the second part of the scheme (8) we take

$$q_\nu = \mu; \quad b_{m\nu} = \begin{cases} 1 & \text{for} \quad 1 \leqslant \nu \leqslant r, \\ 0 & \text{for} \quad \nu > r. \end{cases}$$

Moreover, it is seen that Assumptions 1–5 (cf. §§ 2 and 5 of Part I) are fulfilled in this case.

Like the process studied in the preceding section, this hierarchical process is of Type A. Using formula (9) in Part I, we conclude that the equilibrium probabilities $P_{ij\ldots z}$ are obtained by inserting the following values in the formula (5) of Part I

$$a_{m-1} = 1; \quad b_m = \left.\begin{cases} m & \text{for} \quad m \leqslant r, \\ r & \text{for} \quad m > r, \end{cases}\right\} \quad p_\nu = \lambda_\nu; \quad q_\nu = \mu.$$

When the steady state probabilities have been determined by the above procedure, expressions may be derived for the fractions of time during which the different machines are being serviced, are in working state, etc. In this way generalizations are obtained of the formulae derived in the previous section.

4. Some remarks on inhomogeneous telephone traffic

Palm has written a paper (1937), which has some connexion with the present work. In the last section of the paper, the following telephone trunking problem is considered. A group of n channels is used by N subscribers. Each subscriber has full availability to each channel. The arrangement is supposed to be a busy-signal one, i.e. if all channels are engaged, the subscriber does not wait in a queue, but may make another attempt later on. The traffic is inhomogeneous, i.e. some subscribers use their telephones more frequently than others. It is required to determine the equilibrium probabilities of various states, for example, of the state that all channels are used at the same time.

When solving this problem, Palm used the following assumptions (the notation is ours):

(1) The probability is $\lambda_\nu \Delta t + o(\Delta t)$ that the νth subscriber makes an attempt between t and $t+\Delta t$.

(2) The duration of the call of the νth subscriber follows the mixed-exponential distribution, i.e. the probability of a duration less than a positive quantity t is

$$1 - \sum_{r=1}^{k_\nu} c_{\nu_r} e^{-\mu_{\nu_r} t}, \tag{9}$$

where k_ν is an integer, the quantities μ_{ν_r} are positive and $\sum_{r=1}^{k_\nu} c_{\nu_r} = 1$. Note that the mean duration is equal to $1/\mu_\nu$, where

$$\frac{1}{\mu_\nu} = \sum_{r=1}^{k_\nu} \frac{c_{\nu_r}}{\mu_{\nu_r}}. \tag{10}$$

The purpose of this section is to demonstrate that, in our terminology, Palm studied a hierarchical birth and death process and that his results may be obtained by means of the theory developed in Part I. To see this, we proceed in two steps and begin with a somewhat simpler problem.

(a) *Duration of call follows the exponential distribution*

We then have the following situation. There are N subscribers and n channels ($N > n$). The probability is $\lambda_\nu \Delta t + o(\Delta t)$ that the νth subscriber makes an attempt between t and $t+\Delta t$. Moreover, if a call from this subscriber is going on at time t, the probability is $\mu_\nu \Delta t + o(\Delta t)$ that it ends between t and $t+\Delta t$.

Note that this problem is related to the machine interference problem discussed in § 2 of this paper. Indeed, machines correspond to subscribers, and operators to channels. There are, however, two important differences. First, no queues build up in the present case. Secondly, the probability μ_ν is allowed to depend on ν, which was not the case in the queueing problem.

To solve this problem, it will be convenient to assign numbers $1, \ldots, n$ to the channels in any order and to suppose that, when a call is to begin, a free channel is chosen at random. Denote by $E_{ij\ldots z}$ the state 'the ith, jth, ..., zth subscribers are engaged by calls' and by $E'_{ij\ldots z}$ the substate 'the ith, jth, ..., zth subscribers are engaged by calls and are using channels with numbers in ascending order'. If $P_{ij\ldots z}$ and $P'_{ij\ldots z}$ are the corresponding steady state probabilities, it is concluded that

$$P_{ij\ldots z} = m!\, P'_{ij\ldots z}. \tag{11}$$

The reason why the states $E'_{ij\ldots z}$ are introduced is that transitions between these states follow the scheme (8) of the hierarchical process defined in § 5 of Part I. It is realized that we should take in this scheme

$$a_{m-1,\nu} = \frac{1}{m} \quad (\nu = 1, \ldots, m); \qquad b_{m\nu} = 1 \quad (\nu = 1, \ldots, m); \qquad p_\nu = \lambda_\nu; \qquad q_\nu = \mu_\nu.$$

It is also seen that the process satisfies Assumptions 1–5 (cf. §§ 2 and 5 of Part I) and is of Type B. By formula (9) of Part I we have $a_{m-1} = 1$, $b_m = m$, and it is inferred from formula (5) of Part I that

$$P'_{ij\ldots z} = \frac{1}{m!} \frac{\lambda_i \ldots \lambda_z}{\mu_i \ldots \mu_z} P_0.$$

Using (11) and the notation

$$\rho_\nu = \frac{\lambda_\nu}{\mu_\nu}$$

we find the simple result

$$P_{ij\dots z} = \rho_i \rho_j \cdots \rho_z P_0, \tag{12}$$

where

$$\frac{1}{P_0} = 1 + \Sigma\rho_1 + \Sigma\rho_1\rho_2 + \dots + \Sigma\rho_1\rho_2\cdots\rho_n. \tag{13}$$

The summation in the last term should be made over the $\binom{N}{n}$ terms obtained by taking n indices at a time from the sequence $1, \dots, N$ (and correspondingly for the other terms).

This solution should be compared to the solution (4) and (5) of the machine interference problem in § 2.

(*b*) *Duration of call follows the mixed exponential distribution*

If the duration of a call is distributed according to (9), the following artifice is used. Replace the νth subscriber by a group of k_ν 'sub-subscribers', which are designated by numbers ν_r ($r = 1, \dots, k_\nu$). The probability that the ν_rth sub-subscriber begins a call between t and $t+\Delta t$ is $\lambda_{\nu_r}\,\Delta t + o(\Delta t)$, where

$$\lambda_{\nu_r} = \lambda_\nu c_{\nu_r},$$

and the probability that a call from this sub-subscriber ends between t and $t+\Delta t$ is $\mu_{\nu_r} t + o(\Delta t)$. We also suppose that the sub-subscribers in a group use the same line; i.e. only one of them can use the telephone at a specific moment. It is not difficult to see that these assumptions are entirely equivalent to those used by Palm and quoted at the beginning of this section.

The steady state probability of the event 'the i_αth, j_βth, ..., z_ωth sub-subscribers are engaged by calls' is obtained by taking in the expression (12) the particular values $i = i_\alpha$, $j = j_\beta$, ..., $z = z_\omega$; i.e. we find

$$P_{i_\alpha j_\beta \dots z_\omega} = \rho_{i_\alpha}\rho_{j_\beta}\cdots\rho_{z_\omega}P_0. \tag{14}$$

Adding over all sub-subscribers in the ith, jth, ..., zth groups, we have the simple result

$$P_{ij\dots z} = \rho_i \rho_j \cdots \rho_z P_0,$$

where

$$\rho_\nu = \sum_{r=1}^{k_\nu} \rho_{\nu_r} = \lambda_\nu \sum_{r=1}^{k_\nu} \frac{c_{\nu_r}}{\mu_{\nu_r}}$$

or by (10) more simply

$$\rho_\nu = \frac{\lambda_\nu}{\mu_\nu}.$$

As a consequence, formula (12) holds good not only when the duration is exponentially distributed but, more generally, when it is distributed according to the mixed exponential distribution. This remarkable result is due to Palm. It would be interesting to know if the same result is valid for more general distributions; however, this seems to be an open question.

References

Benson, F. & Cox, D. R. (1951). The productivity of machines requiring attention at random intervals. *J.R. Statist. Soc.* B, **13**, 65.

Blom, G. (1960). Hierarchical birth and death processes. I. Theory. *Biometrika*, **47**, 235.

Feller, W. (1950). *An Introduction to Probability Theory and its Applications*, **1**. New York: John Wiley and Sons.

Palm, C. (1937). Inhomogeneous telephone traffic in full-availability groups. *Ericsson Technics*, no. 1.

Palm, C. (1947). The distribution of repairmen in servicing automatic machines (in Swedish). *Industritidningen Norden*, **75**, 75, 90, 119.

Biometrika (1960), **47**, 3 *and* 4, *p.* 253
Printed in Great Britain

A comparison of the effectiveness of tournaments*

BY W. A. GLENN
Virginia Polytechnic Institute

1. INTRODUCTION AND SUMMARY

When it is desired to compare t treatments in a paired-comparison experiment the simplest balanced design requires $(t-1)$ units of each treatment. Each replication, comprising $\frac{1}{2}t(t-1)$ comparisons, is analogous to a round robin tournament involving t players, with each comparison corresponding to a game between two contestants. This analogy was first pointed out by Kendall (1955). If we are primarily concerned with picking the best treatment it may be practicable to use a technique requiring less than $(t-1)$ units of each treatment per replication. The knock-out tournament (or cup-tie procedure) has been suggested for this purpose by Maurice (1958) and also by David (1959). Maurice compares the balanced design or round robin tournament with the cup-tie procedure for the case of four populations with means in the most unfavourable configuration (one population has a mean larger by δ than the common mean of the other three populations). She bases the comparison on the amount of replication (and hence the sampling cost) required to ensure detection of the difference δ with probability at least P, and concludes that the cup-tie procedure is the more economical. In effecting comparisons with the balanced design she uses tables provided by Bechhofer (1954). David investigates some properties of knock-out and round robin tournaments and obtains an expression for the probability with which the best player wins a knock-out tournament, certain assumptions having been made about the strengths of the players.

In this paper six tournament types are investigated for their effectiveness in selecting the best one of four players. The types, together with abbreviations used for convenience in referring to them, are as follows: R = round robin tournament, K_0 = single knock-out tournament with a random draw and each pair playing one game, K_1 = single knock-out tournament with a random draw and each pair playing until one of the players has won two games, K_2 = combination of two knock-out tournaments with separate random draws, K_3 = combination of two knock-out tournaments with seeding in the second draw based on the outcome of the first tournament, and D = double elimination tournament, in which no player is eliminated until he has lost two games. The K_0 type requires only three games, whereas the other types each require a minimum of six games.

Representing the four players by C_1, C_2, C_3 and C_4, we write π_{ij} for the probability with which C_i defeats C_j. We assume that every game results in the selection of a winner, so that

$$\pi_{ji} = 1 - \pi_{ij}.$$

In § 2 we describe the tournament types and discuss the evaluation of the probability that a specified player emerges as winner, after a play-off if necessary, in each type. Without loss of generality we take C_1 as the specified player. The evaluation of the expected number of games in each type is discussed in § 3. In § 4 a best player is defined and certain assumptions

* Research supported, in part, by the Office of Ordnance Research, U.S. Army.

are made regarding the parameters π_{ij}. Two criteria are used for the comparison of the effectiveness of the tournament types, viz. (*a*) the probability that the best player wins, and (*b*) the expected number of games. Comparisons are based on three series of assigned parameter values. The special case in which all but one of the players are of equal strength (counterpart of the most unfavourable configuration discussed by Maurice) is considered in § 5.

The comparisons in §§ 4 and 5 indicate that if sufficient emphasis is placed on criterion (*a*), K_1 is the best of the six procedures. Otherwise K_0 is preferable. This conclusion is in general agreement with that obtained by Maurice.

2. Evaluation of the probability that C_1 wins

2·1. *Round robin tournament*

In the round robin tournament every player meets every other player once. For four players this yields 64 possible outcomes which may be classified in terms of the partitions

$$[3210], \quad [31^3], \quad [2^30], \quad [2^21^2],$$

occurring respectively in the following numbers of different ways:

$$24, \quad 8, \quad 8, \quad 24.$$

The notation [3210] is used to indicate that one player has won three games, another two, a third one, and the fourth none. The other partitions are similarly defined. Outcomes related to the players in specified order will be enclosed in parentheses. Thus (3210) will represent the outcome in which C_1 has won three games, C_2 has won two games, and so on.

In evaluating the probability that C_1 emerges as winner, one must consider, in addition to the probability of an outright win, the probabilities that C_1 ties with one or two other players and wins the resulting play-off. The probability that C_1 wins outright is readily seen from the outcomes (3210) and (31^3) to be given by $\pi_{12}\pi_{13}\pi_{14}$.

The outcomes in which C_1 ties with two other players are (2220), (2202), and (2022). In writing down the associated probabilities it is helpful to note that the player with no wins has lost to each of the others, while the latter have each won once in games among themselves. Thus, for example, we have

$$P(2220) = \pi_{14}\pi_{24}\pi_{34}(\pi_{12}\pi_{23}\pi_{31} + \pi_{13}\pi_{21}\pi_{32}).$$

The play-off for outcomes in this category takes the form of a round robin tournament with three players, with the possible outcomes [210] and $[1^3]$. If $[1^3]$ occurs the play-off must be repeated. Thus the probability that a given player wins the play-off is conditional upon the non-occurrence of $[1^3]$. For example, the probability that C_1 wins a play-off with C_2 and C_3 may be written as

$$P_{123} = \pi_{12}\pi_{13}/(\pi_{12}\pi_{13} + \pi_{21}\pi_{23} + \pi_{31}\pi_{32}).$$

The complete probability that C_1 ties with two other players and wins the resulting play-off is evaluated from the expression

$$P(2220)\,P_{123} + P(2202)\,P_{124} + P(2022)\,P_{134}.$$

Each of the outcomes (2211), (2121), and (2112), in which C_1 ties with one other player, may occur in four ways, the respective probabilities of which are easily written down. The

complete probability that C_1 ties with one other player and wins the resulting play-off is evaluated from the expression

$$P(2211)\,\pi_{12}+P(2121)\,\pi_{13}+P(2112)\,\pi_{14}.$$

2·2. *Single knock-out tournament* (K_0 *type*)

We designate by K_0 the single knock-out tournament involving four players and consisting of three games, or two rounds of play. For the first round the players are selected into pairs at random. The two winners in the first round play off in the second, so that the final winner will have won two games, another player will have won one, and the other two none. The draw, or selection of the pairs for the first round, may be made in three ways, each of which leads to eight possible outcomes, or 24 outcomes in all. Each of these is the partition $[210^2]$. There are 12 possible outcomes associated with players in specified order, each of which may occur in two ways, i.e. in two out of the three possible draws. Since the three draws are equally likely, the probability (before the draw) that C_i wins two games and C_j wins one in the K_0 tournament is given by the general relation

$$\pi_{ij}(\pi_{ik}\pi_{jm}+\pi_{im}\pi_{jk})/3,$$

where C_k and C_m are the two players with no wins. The probabilities of the 12 outcomes are evaluated from this relation by assigning all possible ordered pairs obtainable from the numbers 1, 2, 3 and 4 to the ordered pair (i, j), and in each case letting k and m be the two numbers not in the pair.

2·3. *Single knock-out tournament* (K_1 *type*)

One way of introducing replication into a single knock-out tournament is to have each pair play more than one game before going on to the next round. We designate by K_1 the single knock-out tournament in which each pair plays until one of the players has won two games (a 'best two out of a possible three' criterion). K_1 is effectively the simple tournament with three games played in each contest, the third being unnecessary when the outcome is determined by the first two. By specifying an odd number of games one takes care of the problem of ties in the individual contests.

The probability that C_i wins two out of a possible three games with C_j is readily seen to be given by

$$Q_{ij} = \pi_{ij}^2(1+2\pi_{ji}).$$

The probabilities associated with the various outcomes in the K_1 tournament are evaluated by following the procedure described in § 2·2 except that all π_{ij}, $i, j = 1, 2, 3, 4$ are replaced by the corresponding Q_{ij}.

2·4. *Combination of two knock-out tournaments* (K_2 *type*)

Replication may also be introduced by repeating the entire tournament, the winner being declared as the player with the largest total number of wins, possibly after a play-off. The simplest way of repeating the knock-out tournament is to start with a fresh random draw for each repetition. We designate by K_2 the combination of two knock-out tournaments based on separate random draws. The two tournaments are evidently independent events. The outcomes of the combination may be classified in terms of the partitions $[420^2]$, $[41^20]$, $[3210]$, $[3^20^2]$, $[2^21^2]$, $[2^30]$, the first three of which correspond to an outright win for one of

the players, while the next two represent ties involving two players, and the last one represents a tie involving three players.

The probabilities associated with the various outcomes may be written by appropriately associating the expressions related to the single tournament described in § 2·2. Play-offs for the ties involving two and three players are, of course, the same as those described in § 2·1.

2·5. *Combination of two knock-out tournaments (K_3 type)*

Instead of using separate random draws, if the tournaments are played in succession, one may use the outcome of the first in setting up the draw for the second. At the end of the first tournament some evidence has been provided as to the relative strengths of the players. The two who have won games appear to be stronger than the two with no wins. We wish to investigate the possibility that this information may be used to improve the effectiveness of the combination of two tournaments.

In the K_3 type we select the draw for the second tournament so that the two players with wins in the first tournament remain in opposite halves while the positions of their opponent e interchanged. The first tournament is based on a random draw, and its outcome together with the particular draw used uniquely determine the draw for the second tournament. We wish to keep the two players who have appeared best in the first tournament apart in the first round of the second. Only two of the three possible draws will accomplish this, and the arrangement which we have proposed selects the one of these which was not used in the first tournament. Thus the only alternative seeding arrangement would involve repeating the first draw. Further reference to this point will be made in § 4.

The possible outcomes of the K_3 combination are the same as those of K_2, but the tournaments are not independent events. The conditional probabilities involved are obtained by appropriate association of the probabilities of the outcomes of the K_0 tournament together with the draws from which they arise.

2·6. *Double elimination tournament*

In the double elimination tournament no player is eliminated until he has lost two games. The first round is the same as for the single knock-out tournament (K_0), but in the second round, in addition to the two winners, the two losers are entered in play. Of these, one will be a two-time loser at the end of the second round, and is then eliminated from further play. This leaves one two-time winner and two with a win and a loss, which we shall designate respectively by WW, WL, and LW, letting W stand for a win and L for a loss. In the fifth game WL and LW play off for the elimination of the loser. The winner, either WLW or LWW, meets WW in the sixth game. Should WW win, his opponent is eliminated and the tournament ends, the winner being represented by WWW. However, should WW lose, this being his first loss, he then has the record WWL whereas his opponent in the sixth game now has the record $WLWW$ or $LWWW$ as the case may be. A seventh game is then required to determine the final winner, represented by $WWLW$, $WLWWW$, or $LWWWW$ as the case may be.

The probability that C_1 wins is the sum of the probabilities that he achieves one of the winning outcomes WWW, $WWLW$, $WLWWW$, or $LWWWW$. This is easily evaluated for a given draw and the result for each of the other two possible draws may be inferred by appropriate interchanges of subscripts. For the random draw the resulting values are averaged over the three possible draws.

3. Evaluation of the expected number of games

3·1. *Round robin tournament*

The round robin tournament with four players ends in six games only if one player achieves an outright win. Play-offs involving three players require initially three additional games, three more if $[1^3]$ occurs, and so on, leading to an infinite geometric progression with common ratio $P(1^3)$. Thus, for example, the expected number of additional games in a play-off involving C_1, C_2, and C_3 may be written as

$$E_{123} = 3/(\pi_{12}\pi_{13}+\pi_{21}\pi_{23}+\pi_{31}\pi_{32}).$$

The expected number of additional games due to ties involving three players is evaluated from the expression

$$P(2220)\,E_{123}+P(2202)\,E_{124}+P(2022)\,E_{134}+P(0222)\,E_{234}.$$

Each of the outcomes in the classification $[2^21^2]$ leads to a single-game play-off. Thus the expected number of additional games due to ties involving two players is the sum of the probabilities of the outcomes (2211), (2121), (2112), (1221), (1212), and (1122).

3·2. *Single knock-out tournaments (K_0 and K_1 types)*

The K_0 tournament will always consist of three games, since no tie-breaking play-offs are required. However, the K_1 type requires a minimum of six games, since at least two are needed in each of the three contests. A third game is required whenever one player wins the first game and the other the second. Thus in a given contest the probability of a third game is the probability that a given player wins the first game and loses the second, or vice-versa. For a given draw the excess over six of the expected number of games is the sum of the probabilities of third games in all contests evaluated for all possible outcomes. For the random draw these values are averaged over the three possible draws.

3·3. *Combinations of two knock-out tournaments (K_2 and K_3 types)*

In the combinations of two knock-out tournaments if any player achieves an outright win the series ends in six games. Ties involving two players and represented by the partitions $[3^20^2]$ and $[2^21^2]$ each require a single-game play-off. The expected number of additional games on account of these is therefore the sum of the probabilities of the appropriate outcomes. The number of additional games due to ties involving three players is evaluated in the manner described in § 3·1.

3·4. *Double elimination tournament*

The minimum number of games required in the double elimination tournament with four players is six, since three players must lose two games each. Thus the expected number of games is six plus the probability of a seventh game. Such a game is required when WW loses in the sixth game. Using the record of C_1 as a basis of classification we may say that a seventh game is required in the event of each of the following outcomes: $WWLW$, $WWLL$, $WLWWW$, $WLWWL$, $WWWW$, and $LWWWL$. Since the probability of a seventh game is not affected by the winning or losing of C_1 in that game, we may group these outcomes in pairs to give WWL, $WLWW$, and $LWWW$. In addition, some of the outcomes in which C_1 is eliminated earlier in the tournament may lead to a seventh game. They are WLL, LWL, and LL, a

seventh game being required in these cases whenever the player with no previous losses loses in the sixth game. Using this classification system, one may determine the probability of a seventh game for a given draw. The result for each of the other two possible draws may be inferred by appropriate interchanges of subscripts. For the random draw the resulting values are averaged over the three possible draws.

4. COMPARISONS IN SERIES OF NUMERICAL EXAMPLES

For the purpose of comparing the effectiveness of the six tournament types described above, we define a best player as one having a probability greater than $\frac{1}{2}$ of defeating each of the others. Without loss of generality we may refer to the players in decreasing order of strength, C_1 being strongest and C_4 being weakest. The possibility that some of the players C_2, C_3 and C_4 may be equally strong is not excluded. We shall consider only situations in which circular triads amongst the parameters do not occur; accordingly we assume that the six parameters satisfy the following inequalities:

$$\pi_{14} \geqslant \pi_{13} \geqslant \pi_{12} > \tfrac{1}{2}, \quad \pi_{24} \geqslant \pi_{23}' \geqslant \tfrac{1}{2}, \quad \pi_{13} \geqslant \pi_{23} \geqslant \tfrac{1}{2}, \quad \pi_{14} \geqslant \pi_{24} \geqslant \pi_{34} \geqslant \tfrac{1}{2}. \tag{4·1}$$

Table 4·1. *Assigned parameter values in three series of examples*

	Example no.	π_{23}	π_{24}	π_{34}
Series 1	1·1	0·5000	0·5000	0·5000
$\pi_{12} = 0{\cdot}5400$	1·2	·6400	·8500	·6200
$\pi_{13} =$ ·6500	1·3	·6127	·8396	·7679
$\pi_{14} =$ ·8600	1·4	·6400	·8500	·8200
	1·5	·6500	·8600	·8400
Series 2	2·1	0·5000	0·5000	0·5000
$\pi_{12} = 0{\cdot}7000$	2·2	·5758	·7247	·6598
$\pi_{13} =$ ·7600	2·3	·7500	·8200	·7200
$\pi_{14} =$ ·8600	2·4	·7500	·8400	·8000
	2·5	·7500	·8500	·8400
Series 3	3·1	0·5000	0·5000	0·5000
$\pi_{12} = 0{\cdot}8000$	3·2	·5862	·6923	·6136
$\pi_{13} =$ ·8500	3·3	·7000	·7500	·7000
$\pi_{14} =$ ·9000	3·4	·8000	·8300	·7500
	3·5	·8400	·8700	·8500

Comparisons have been made by assigning values to the six parameters and evaluating the two criteria, viz. (*a*) the probability that C_1 wins, and (*b*) the expected number of games for each of the tournament types. Typical results in the form of three series of examples are presented below. The assigned parameter values, satisfying the inequalities (4·1) are given in Table 4·1. In the examples 1·3, 2·2, and 3·2 the values of π_{23}, π_{24}, and π_{34} have been determined from the assigned values of π_{12}, π_{13}, and π_{14} so as to satisfy the mathematical model proposed by Bradley & Terry (1952). For the three series and each of the six tournament types the probabilities that C_1 wins are given in Table 4·2, listed in generally increasing order of the probabilities involved. The same order of presentation is used in Table 4·3, which lists the expected numbers of games for all types except K_0 (which always requires three games).

It will be observed in Table 4·1 that each series is characterized by a common set of values of π_{12}, π_{13}, and π_{14} in the five examples. Within series the values of π_{23}, π_{24}, and π_{34} have been

assigned in a generally increasing order of magnitude in the ranges defined by (4·1), starting in each case with the lower limit of $\frac{1}{2}$. With this arrangement one finds that the probability that C_1 wins decreases throughout each series for all six tournament types. This is because we are giving an increasing advantage to C_2, and to some extent to C_3, relative to the weakest player C_4, thus tending to increase the probability that C_1 will meet later in the tournament or in a play-off a player that he has the relatively smaller chance of defeating.

Table 4·2. *Probabilities that C_1 wins in three series of examples*

Example no.	Tournament type					
	K_0	R	K_2	K_3	D	K_1
1·1	0·4581	0·5239	0·5279	0·5284	0·5307	0·5374
1·2	·4249	·4615	·4641	·4669	·4711	·4789
1·3	·4209	·4564	·4581	·4621	·4672	·4737
1·4	·4173	·4503	·4502	·4551	·4618	·4688
1·5	·4156	·4472	·4463	·4516	·4593	·4667
2·1	0·5958	0·6910	0·6974	0·7036	0·7212	0·7406
2·2	·5817	·6637	·6714	·6795	·6992	·7179
2·3	·5734	·6438	·6452	·6564	·6860	·7063
2·4	·5707	·6393	·6390	·6514	·6826	·7034
2·5	·5694	·6370	·6357	·6488	·6810	·7022
3·1	0·7216	0·8190	0·8255	0·8341	0·8592	0·8751
3·2	·7133	·8028	·8119	·8246	·8491	·8652
3·3	·7089	·7924	·7996	·8115	·8436	·8602
3·4	·7044	·7803	·7832	·7977	·8381	·8562
3·5	·7014	·7734	·7726	·7894	·8347	·8540

Table 4·3. *Expected numbers of games in three series of examples*

Example no.	Tournament type				
	R	K_2	K_3	D	K_1
1·1	6·8175	6·7796	6·7630	6·4543	7·3301
1·2	6·7320	6·6978	6·6780	6·4761	7·2401
1·3	6·7740	6·7174	6·6908	6·4812	7·2255
1·4	6·7757	6·7121	6·6831	6·4818	7·1962
1·5	6·7769	6·7097	6·6794	6·4811	7·1808
2·1	6·6588	6·6298	6·6215	6·4072	7·2080
2·2	6·6406	6·6139	6·6026	6·4250	7·1741
2·3	6·6197	6·5699	6·5649	6·4267	7·0836
2·4	6·6298	6·5724	6·5629	6·4280	7·0502
2·5	6·6355	6·5745	6·5620	6·4286	7·0298
3·1	6·5120	6·4496	6·4514	6·3218	7·0202
3·2	6·4683	6·4453	6·4569	6·3361	6·9984
3·3	6·4611	6·4331	6·4337	6·3419	6·9468
3·4	6·4523	6·4109	6·4154	6·3458	6·8685
3·5	6·4547	6·4048	6·4055	6·3473	6·7936

In Table 4·2 we observe that K_2 is slightly better than R in the early part of each series, but that this effect tends to be reversed for the larger values of π_{23}, π_{24}, and π_{34}. The seeding arrangement (K_3) produces only a small improvement over the use of separate random draws (K_2). It was remarked in § 2·5 that the only alternative seeding arrangement would involve

making a single random draw and using it for both tournaments. This has been found in a number of examples to give results which are intermediate between those for K_2 and K_3. The latter thus appears to be the optimum available seeding arrangement. In terms of the probability that C_1 wins, K_0 is least desirable, D is somewhat better than R, K_2, and K_3, but K_1 is superior to all of the others.

In Table 4·3 we observe that the expected number of games is less for the double elimination tournament than for each of the R, K_2, and K_3. Thus, for the cases considered, the round robin and the two knock-out combinations K_2 and K_3 always compare unfavourably with the double elimination procedure in terms of both criteria. However, the K_1 type apparently accomplishes a higher probability that C_1 wins at the expense of a higher expected number of games, while K_0, although yielding the lowest probability, requires only three games.

In order to effect reasonable comparisons between D, K_0, and K_1 we set up the cost function

$$\text{expected cost} = q[1 - P(C_1 \text{ wins})] + \text{expected number of games}, \qquad (4{\cdot}2)$$

where q is a factor representing the assessed cost of a wrong decision, expressed in units of the cost per game.

Using (4·2) one may find for each example three critical values of q, viz.

q_1: above which D is preferable to K_0;
q_2: above which K_1 is preferable to D;
q_3: above which K_1 is preferable to K_0.

If $q_1 < q_2$ the interval between these two values is a range of q for which D is better than K_0 and K_1. However, if $q_1 \geqslant q_2$ there is no value of q for which D is to be preferred to K_0 and K_1.

Table 4·4. *Critical values of q in three series of examples*

Example no.	q_1	q_2	q_3	Example no.	q_1	q_2	q_3	Example no.	q_1	q_2	q_3
1·1	47	130	54	2·1	27	41	29	3·1	24	43	26
1·2	75	97	78	2·2	29	40	30	3·2	24	41	26
1·3	75	114	80	2·3	30	32	30	3·3	24	36	26
1·4	78	102	81	2·4	30	29	30	3·4	25	28	25
1·5	79	94	81	2·5	30	28	30	3·5	25	23	24

For the three series of examples considered in this section the critical values q_1, q_2, and q_3, rounded to the nearest integer below the exact value, are given in Table 4·4. We observe that in Examples 2·4, 2·5, and 3·5 there is no value of q for which D is preferable to K_0 and K_1. In the other cases, in which there is a range of q where D is preferable to K_0 and K_1, that range is observed to be, in general, relatively short. Furthermore, upon evaluating the advantage of using D instead of K_1, for $q = q_3$ and also for some other intermediate values in the range between q_1 and q_2, it has been found that the largest advantage amounts to only 0·51 (Example 1·1). That is, the advantage of using D instead of K_1 in a range in which D is preferable seems to be only about half the cost of one game. In view of these considerations it would seem to be best to use K_0 for $q < q_3$ and K_1 for $q > q_3$. Since K_1 is effectively the simple tournament with replication in each contest, we conclude that it seems to be better to vary the amount of experimentation than to depart from the ordinary (and simplest) form of tournament.

5. Comparisons in a special case

For the special case in which C_2, C_3, and C_4 are equal in strength we may write

$$\pi_{12} = \pi_{13} = \pi_{14} = \pi \quad \text{and} \quad \pi_{23} = \pi_{24} = \pi_{34} = \tfrac{1}{2}. \tag{5·1}$$

Expressions for the probability that C_1 wins and the expected number of games in each of the six tournament types are found in the manner indicated in §§ 2 and 3. If we denote the probability that C_1 wins by P and use the abbreviations for the tournament types as subscripts, we find that

$$\left.\begin{aligned} P_R &= \pi^3(2-\pi)+\pi^3(1-\pi)(2+\pi+2\pi^2)/4(1-\pi+\pi^2),\\ P_{K_0} &= \pi^2,\\ P_{K_1} &= \pi^3(2-\pi)+\pi^3(1-\pi)(-2+8\pi-4\pi^2),\\ P_{K_2} &= \pi^3(2-\pi)+2\pi^4(1-\pi)(3-2\pi+\pi^2)/3(1-\pi+\pi^2),\\ P_{K_3} &= \pi^3(2-\pi)+\pi^4(1-\pi)(4-3\pi+2\pi^2)/2(1-\pi+\pi^2),\\ \text{and}\quad P_D &= \pi^3(2-\pi)+2\pi^4(1-\pi). \end{aligned}\right\} \tag{5·2}$$

Similarly, using E to denote the expected number of games, we have

$$\left.\begin{aligned} E_R &= 7-3\pi/2+3\pi^2/2-\pi^3+9\pi^2(1-\pi)/4(1-\pi+\pi^2),\\ E_{K_0} &= 3,\\ E_{K_1} &= 7+2\pi-7\pi^2/2+7\pi^3-10\pi^4+4\pi^5,\\ E_{K_2} &= 41/6-\pi+\pi^2/6+2\pi^3/3-2\pi^4/3+2\pi^2(1-\pi)(2-\pi)/(1-\pi+\pi^2),\\ E_{K_3} &= 27/4-\pi/2-\pi^2/4+\pi^3-\pi^4+3\pi^2(1-\pi)(2-\pi)/2(1-\pi+\pi^2),\\ \text{and}\quad E_D &= 13/2-\pi^2/2+2\pi^3-2\pi^4. \end{aligned}\right\} \tag{5·3}$$

For the special case, the inequalities (4·1) reduce to

$$1 > \pi > \tfrac{1}{2}. \tag{5·4}$$

From the expressions (5·2) together with the condition (5·4) it may be shown that

$$P_{K_1} > P_D > P_{K_3} > P_{K_2} > P_R > P_{K_0}. \tag{5·5}$$

Since it may be shown that E_D is less than each of E_R, E_{K_1}, E_{K_2}, and E_{K_3}, the double elimination tournament is preferable to each of R, K_2, and K_3 in terms of both criteria. However, as in § 4, it may be observed that K_1 yields the highest probability that C_1 wins at the expense of the highest expected number of games, while K_0, although giving the lowest probability, requires only three games.

For the comparison of D, K_0, and K_1 we again use the cost function (4·2). The critical values q_1, q_2, and q_3 have been evaluated for nine values of π, uniformly spaced over the range defined in (5·4), and are given, rounded to the nearest integer below the exact value, in Table 5·1. Inspection of these values indicates that for $\pi < 0{\cdot}70$ there is no range of values of q for which D is preferable to K_0 and K_1. For $\pi \geqslant 0{\cdot}75$ there is such a range, the length of which tends to increase as π increases. However, upon evaluating the advantage of using D instead of K_1 for $\pi = 0{\cdot}95$ and $q = q_3$, it has been found that the advantage is only about 0·40, or less than half the cost of one game. For smaller values of π the advantage is less than this. Thus we have again the conclusion that it would seem to be best to use K_0 for $q < q_3$

and K_1 for $q > q_3$, the choice therefore depending on the emphasis to be placed on the criterion of reliability relative to that of sampling cost. It would appear that, at least for four players, there is not much to be gained from varying the ordinary tournament rules, so long as one allows variation in the specified number of wins.

Table 5·1. *Critical values of q in the special case*

π	q_1	q_2	q_3
0·55	165	147	161
·60	75	70	74
·65	47	46	47
·70	34	36	34
·75	27	31	28
·80	24	31	25
·85	23	36	25
·90	26	53	28
·95	39	144	44

The author is indebted to Dr H. A. David for suggesting the topic of this research. Appreciation is also expressed to a referee who made some valuable suggestions concerning the original manuscript.

REFERENCES

BECHHOFFER, R. E. (1954). A single-sample multiple decision procedure for ranking means of normal populations with known variances. *Ann. Math. Statist.* **25**, 16–39.

BRADLEY, R. A. & TERRY, M. E. (1952). Rank analysis of incomplete block designs. I. The method of paired comparisons. *Biometrika*, **39**, 324–45.

DAVID, H. A. (1959). Tournaments and paired comparisons. *Biometrika*, **46**, 139–49.

KENDALL, M. G. (1955). Further contributions to the theory of paired comparisons. *Biometrics*, **11**, 43–62.

MAURICE, RITA J. (1958). Selection of the population with the largest mean when comparisons can be made only in pairs. *Biometrika*, **45**, 581–6.

Biometrika (1960), **47**, 3 *and* 4, *p.* 263
Printed in Great Britain

Supplemented balance

BY S. C. PEARCE
East Malling Research Station

1. INTRODUCTION

In comparative experiments there may be a treatment, usually the control, which is logically in a different position from the rest. For example, the purpose of the trial may be to compare some new treatments not with each other but with an established one to find out if any are worth further study; in this instance it is known that the greatest economy is obtained if the control is more highly replicated than the others by a ratio equal to the square root of the number of new treatments under test. On the other hand, a trial may be designed to suggest which of several courses of action is best and the control is introduced only in order to demonstrate the bad consequences of doing nothing. In such circumstances a single plot may suffice.

When designing such a trial an experimenter may well use an orthogonal design as being more convenient even though it may not give the optimum replication to the treatment in a special position. However, if some restriction on block size or plot number constrains him to use a non-orthogonal design, supplemented balance may be of use to him.

This device was suggested by Hoblyn, Pearce & Freeman (1954) in another connexion, namely, the changing of treatments in successive experiments in cases where it can be assumed that the new treatments will not interact with residual effects from those previously applied. According to their definition a design is in supplemented balance if all treatments are replicated r times except the supplementing one, which has r_0 replicates, and if all pairs of treatments concur λ times in blocks, unless one of the pair is the supplementing treatment, in which case there are λ_0 concurrences. It is assumed that all blocks contain k plots. The treatments will be numbered 0 to v, treatment 0 being the supplementing one; the number of blocks will be written b.

Expressed more formally, let $\mathbf{n}$ be an incidence matrix having b columns and $(v+1)$ rows, the elements representing the number of times each treatment occurs in each block. Also, let $\mathbf{1}_h$ be a unit column vector with h elements, and let $\mathbf{r}$ be a column vector formed from the row totals of $\mathbf{n}$ and thus representing the numbers of replicates of the treatments ($\mathbf{r} = \mathbf{n}\mathbf{1}_b$). Similarly let $\mathbf{k}$ be another column vector formed from the column totals of $\mathbf{n}$ thus representing the numbers of plots in the blocks ($\mathbf{k} = k\mathbf{1}_b$, $\mathbf{k} = \mathbf{n}'\mathbf{1}_{(v+1)}$).

Let $\mathbf{n}_j$ be the jth row of $\mathbf{n}$, then $\lambda = \mathbf{n}_i\mathbf{n}_j'$ if neither i nor j equals 0, and $\lambda_0 = \mathbf{n}_0\mathbf{n}_j'$. Further, let $s_j = \mathbf{n}_j\mathbf{n}_j'$. This quantity may be termed the number of self-concurrences of treatment j, and equals the number of replicates if all elements of $\mathbf{n}_j$ are either 0 or 1.

Considering now the matrix, $\mathbf{n}\mathbf{n}'$, its diagonal elements will have the values s_j, and its off-diagonal elements will equal λ, unless the row or column refers to the supplementing treatment in which case the element equals λ_0. Now

$$k\mathbf{r} = k\mathbf{n}\mathbf{1}_b = \mathbf{n}\mathbf{k} = \mathbf{n}\mathbf{n}'\mathbf{1}_{(v+1)}.$$

Hence
$$kr = s_j + (v-1)\lambda + \lambda_0, \quad \text{where} \quad j \neq 0,$$

so s_j must be independent of j and equal to s, say. Also

$$kr_0 = s_0 + v\lambda_0.$$

A design will here be defined as being of type S if all blocks are the same size and **nn'** has the pattern

$$\begin{array}{ccccc} s_0 & \lambda_0 & \lambda_0 & \dots & \lambda_0 \\ \lambda_0 & s & \lambda & \dots & \lambda \\ \lambda_0 & \lambda & s & \dots & \lambda \\ \vdots & \vdots & \vdots & & \vdots \\ \lambda_0 & \lambda & \lambda & \dots & s \end{array}$$

In this paper reference will also be made to designs of types O, T and P. These have been defined previously (Hoblyn *et al.* 1954) but they can readily be redefined in the present notation. Those of type O are orthogonal, and Tocher (1952) has shown that for such designs all the columns of **n** must be the same, if all blocks are the same size. Consequently if all treatments are equally replicated, all the elements of **nn'** will have the same value, s. In designs of types T and P all blocks are of equal size and all treatments equally replicated. In type T the incidence matrix, **n**, is such that all diagonal elements of **nn'** have the value s, and all off-diagonal elements the value λ. This type thus includes all the balanced incomplete block designs. In type P any row or column of **nn'** can be obtained from any other row or column by permutation, all the diagonal elements being equal to s. Thus all partially balanced incomplete block designs are of type P.

By a design of type X : YZ will be meant one in which there are three classifications of the plots, the second classification being disposed with regard to the first in a design of type X, the third to the first in a design of type Y and the third to the second in a design of type Z. This notation can readily be extended if there are more classifications.

2. Examples of designs of type S

An example is afforded by the strawberry weedkiller trial described by Pearce (1953, § 38). In it there were four blocks ($b = 4$) each of seven plots ($k = 7$), and there were four treatments, A, B, C and D, ($v = 4$) besides an untreated control (O). It was designed thus

Block			
I	II	III	IV
O	O	O	O
O	O	O	O
A	A	A	A
A	B	B	B
B	C	B	C
C	D	C	C
D	D	D	D

Here $r = 5$, $r_0 = 8$, $\lambda = 6$, $\lambda_0 = 10$, $s = 7$ and $s_0 = 16$.

Another example is provided by a strawberry manurial trial recently designed at East Malling. In this the non-supplementing treatments formed a factorial set, being made up of dung (D) or no dung (N) applied before planting, and three levels of inorganic nitrogen

(0, 1 or 2) applied annually. The supplementing treatment (S) was straw incorporated into the soil before planting to depress the nitrogen level. The design was

Block								
I	II	III	IV	V	VI	VII	VIII	IX
N 0	N 0	N 1	N 0	N 0	N 0	N 0	N 0	N 0
N 1	N 1	N 2	N 1	N 1	N 1	N 1	N 1	N 2
N 2	N 2	D 0	N 2	N 2	N 2	D 0	N 2	D 0
D 1	D 0	D 1	D 0	D 0	D 0	D 1	D 0	D 1
D 2	D 2	D 2	D 1	D 1	D 1	D 2	D 1	D 2
S	S	S	D 2	S	D 2	S	D 2	S

Here $r = s = 8$, $r_0 = s_0 = 6$, $\lambda = 7$ and $\lambda_0 = 5$.

Some similar designs were described by Tocher (1952), for example, A and C on his page 85.

3. The analysis of data when the design is of type S

Let α be a general parameter, and let $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ be column vectors of the block and treatment parameters, respectively, subject to the constraints that $\mathbf{k}'\boldsymbol{\beta} = 0$ and $\mathbf{r}'\boldsymbol{\gamma} = 0$.

Also, let $\mathbf{B}$ and $\mathbf{T}$ be column vectors of the block and treatment totals respectively, and let $\mathbf{r}^s$ be $\mathbf{r}$ written as a diagonal matrix.

Then, following the lines indicated by Tocher (1952),

$$\mathbf{T} = \mathbf{r}\alpha + \mathbf{n}\boldsymbol{\beta} + \mathbf{r}^s\boldsymbol{\gamma} + \text{an error term},$$

$$\mathbf{B} = \mathbf{k}\alpha + k\boldsymbol{\beta} + \mathbf{n}'\boldsymbol{\gamma} + \text{an error term}.$$

Defining a column vector, $\mathbf{Q}$, by the equation

$$k\mathbf{Q} = k\mathbf{T} - \mathbf{n}\mathbf{B},$$

then

$$k\mathbf{Q} = (k\mathbf{r}^s - \mathbf{n}\mathbf{n}')\boldsymbol{\gamma} + \text{an error term}.$$

Writing 'the expectation of' as E, it follows that for $j \neq 0$

$$\begin{aligned} E(kQ_j) &= (kr - s)\gamma_j - \lambda_0\gamma_0 - \lambda \sum_{h \neq 0, j} \gamma_h \\ &= (kr - s + \lambda)\gamma_j + (r_0\lambda - r\lambda_0)\gamma_0/r \\ &= k(\phi\gamma_j + \zeta\gamma_0), \quad \text{say}. \end{aligned}$$

$$\begin{aligned} E(kQ_0) &= (kr_0 - s_0)\gamma_0 - \lambda_0 \sum_{h \neq 0} \gamma_h \\ &= (kr_0 - s_0 + r_0\lambda_0/r)\gamma_0 \\ &= k\phi'\gamma_0, \quad \text{say}. \end{aligned}$$

Since

$$kr = s + (v-1)\lambda + \lambda_0, \quad kr_0 = s_0 + v\lambda_0,$$

it follows that

$$\phi = (v\lambda + \lambda_0)/k, \quad \phi' = b\lambda_0/r$$

(N.B. the total number of plots in the experiment, N, is expressible either as bk or as $vr + r_0$.)

The treatment parameters can accordingly be estimated in terms of the data thus:

$$\hat{\gamma}_0 = Q_0/\phi',$$

and, for j not equal to 0, $$\hat{\gamma}_j = (Q_j - \zeta Q_0/\phi')/\phi.$$

Using these estimates it is known from the work of Tocher (1952) that W, the treatment sum of squares, equals $\mathbf{Q}'\hat{\boldsymbol{\gamma}}$. It is most easily computed in that form, but, because the elements of $\mathbf{Q}$ sum to zero, it may also be written

$$\frac{1}{\phi}\sum_{h=1}^{v} Q_h^2 + \frac{\phi+\zeta}{\phi\phi'}Q_0^2 = \frac{1}{\phi}\left\{\sum_{h=1}^{v} Q_h^2 + \frac{\lambda}{\lambda_0}Q_0^2\right\}.$$

Since the sum of squares within blocks is made up of two components, one due to treatments and the other to error, the evaluation of the first of these enables the analysis to be completed readily.

The adjusted mean of the jth treatment equals $(\alpha+\gamma_j)$, α being estimated by the general mean, namely G/N, where G is the grand total of data.

The difference of $\hat{\gamma}_i$ and $\hat{\gamma}_j$, where neither i nor j equals zero, is given by $(Q_i - Q_j)/\phi$; hence it follows from the lemma given by Freeman (1957) that the standard error of the difference of two adjusted means, the supplementing treatment not being involved, is

$$\{2 \times \text{error mean-square}/\phi\}^{\frac{1}{2}}.$$

If, however, the supplementing treatment is being compared with one of the others the standard error is

$$\{(\phi+\phi'+\zeta)\,\text{error mean-square}/(\phi\phi')\}^{\frac{1}{2}} = \{(1+\lambda/\lambda_0)\,\text{error mean-square}/\phi\}^{\frac{1}{2}}.$$

All comparisons will have equal precision if and only if $\lambda = \lambda_0$.

Another special case arises when $r = r_0$. Although supplemented balance has been recommended on account of its permitting unequal replication, the definition does not require this property. Usually in a general three-class design of type X:YZ, if Y = S and the replication is unequal, Z also must equal S. If, however, replication is equal this is not necessary.

4. A NUMERICAL EXAMPLE

Some data from the strawberry weedkiller trial mentioned above have been published (Pearce, 1953, Appendix III). They will here be analysed again using the method just given.

The purpose of the trial was to find out whether any of the weedkillers, A, B, C or D, all of which were apparently suitable for controlling weeds in strawberry fields, would harm the fruiting plants. The data represent the total spread in inches of twelve plants per plot approximately two months after the application of the weedkillers. The comparisons of most interest are those of A, B, C and D with the control, O.

Block I	Block II	Block III	Block IV
C, 107	A, 136	B, 118	O, 173
A, 166	O, 146	A, 117	C, 95
D, 133	C, 104	O, 176	C, 109
B, 166	B, 152	D, 132	A, 130
O, 177	D, 119	B, 139	D, 103
A, 163	O, 164	O, 186	O, 185
O, 190	D, 132	C, 103	B, 147

The incidence matrix, **n**, is

	I	II	III	IV
O	2	2	2	2
A	2	1	1	1
B	1	1	2	1
C	1	1	1	2
D	1	2	1	1

The block totals are respectively 1102, 953, 971 and 942. The grand total, G, is 3968, and the treatment totals respectively 1397, 712, 722, 518 and 619.

The next task is to evaluate the Q's:

$$kQ_O = 7.1397-(2.1102+2.953+2.971+2.942) = +1843,$$
$$kQ_A = 7.712\ -(2.1102+1.953+1.971+1.942) = -\ 86,$$
$$kQ_B = 7.722\ -(1.1102+1.953+2.971+1.942) = +\ 115,$$
$$kQ_C = 7.518\ -(1.1102+1.953+1.971+2.942) = -1284,$$
$$kQ_D = 7.619\ -(1.1102+2.953+1.971+1.942) = -\ 588.$$

These quantities should always sum to zero.

Here it should be noted that

$$k\phi = v\lambda+\lambda_0 = 34, \quad kr\phi' = bk\lambda_0 = 280,$$

and

$$kr\zeta = r_0\lambda - r\lambda_0 = -2.$$

Hence,

$$(k\phi)(kr\phi')\hat{\gamma}_O = r(k\phi)(kQ_O) = 313{,}310,$$
$$(k\phi)(kr\phi')\hat{\gamma}_A = (kr\phi')(kQ_A)-(kr\zeta)(kQ_O) = -\ 20{,}394,$$
$$(k\phi)(kr\phi')\hat{\gamma}_B = (kr\phi')(kQ_B)-(kr\zeta)(kQ_O) = \ 35{,}886,$$
$$(k\phi)(kr\phi')\hat{\gamma}_C = (kr\phi')(kQ_C)-(kr\zeta)(kQ_O) = -355{,}834,$$
$$(k\phi)(kr\phi')\hat{\gamma}_D = (kr\phi')(kQ_D)-(kr\zeta)(kQ_O) = -160{,}954.$$

When weighted by the numbers of replicates these quantities must always sum to zero.

The treatment sum of squares is obtained by multiplying each value of Q by the corresponding value of $\hat{\gamma}$ and adding the products. This gives 1,134,842,912/7.34.280 = 17,029·45, which is nearly the same as the figure of 17,029·49 obtained previously using a different method (Pearce, 1953, Appendix III). The total sum of squares of data within the experiment is 22,855, of which 20,490 lies within blocks and 2,365 between blocks, so the analysis of variance table is:

Source	d.f.	Sum of squares	Mean-square
Blocks (unadjusted)	3	2,365	—
Treatments (adjusted)	4	17,029	—
Error (by difference)	20	3,461	173·0
Total	27	22,855	—

This inquiry specifically concerns the difference between each weedkiller and the control. Formulae have been given for adjusted treatment means, but as long as only differences are in question, it is enough to compare parameters, thus:

$$\hat{\gamma}_O-\hat{\gamma}_A = 35{\cdot}05, \quad \hat{\gamma}_O-\hat{\gamma}_B = 29{\cdot}14,$$
$$\hat{\gamma}_O-\hat{\gamma}_C = 70{\cdot}29, \quad \hat{\gamma}_O-\hat{\gamma}_D = 49{\cdot}82.$$

All these differences are positive, and suggest that all the weedkillers have harmed the strawberry plants to some extent. The standard error of these differences is

$$\{(1+\lambda/\lambda_0)\,\text{error mean-square}/\phi\}^{\frac{1}{2}} = 7{\cdot}55,$$

and this shows that all the differences are in fact significant at a high level, especially when it is considered that one-sided tests may be used.

It may be noted that the efficiency factor for these comparisons is 221/224, or virtually unity.

5. Designs of type O : SS

In these there are initially two mutually orthogonal classifications (type O), whether of rows and columns or of blocks and former treatments, and a third classification (treatments) is made in such a manner as to be in supplemented balance with regard to both the earlier classifications (type O : SS). For example, in these two instances the treatments are so disposed with regard to both rows and columns:

A	A	B	C	O
B	O	C	A	B
C	B	O	O	A
O	C	A	B	C
C	A	B	O	O

A	B	C	O	A
O	C	A	B	C
B	A	O	C	B

In the first example the values of k, λ, s, etc., are the same for both schemes of supplementation; in the second they are different. Both designs belong to special classes already described (Pearce, 1952). Latin squares with a row and column missing (Yates, 1936) are also of type O : SS.

In these designs there are in effect two sets of blocks, whatever names may best describe the two initial classifications in particular cases, and there is a scheme of supplemented balance for each. The same notation will be used as before, one system of blocking being indicated by a cross and the other by a circle. There will be two incidence matrices, $\mathbf{n}^+$ and $\mathbf{n}^\circ$, two vectors of block parameters, $\boldsymbol{\beta}^+$ and $\boldsymbol{\beta}^\circ$, and so on.

For these designs $\mathbf{Q}$ is defined as

$$\mathbf{T} - (\mathbf{n}B/k)^+ - (\mathbf{nB}/k)^\circ + \mathbf{r}G/N = [\mathbf{r}^\delta - (\mathbf{nn}'/k)^+ - (\mathbf{nn}'/k)^\circ]\boldsymbol{\gamma},$$

i.e.
$$\begin{aligned} Q_j (j \neq 0) &= [r - (s/k)^+ - (s/k)^\circ]\gamma_j - [(\lambda/k)^+ + (\lambda/k)^\circ] \sum_{h \neq j,\,0} \gamma_h - [(\lambda_0/k)^+ + (\lambda_0/k)^\circ]\gamma_0 \\ &= \left\{r - \left(\frac{s-\lambda}{k}\right)^+ - \left(\frac{s-\lambda}{k}\right)^\circ\right\}\gamma_j + \left\{\left(\frac{r_0\lambda - r\lambda_0}{kr}\right)^+ + \left(\frac{r_0\lambda - r\lambda_0}{kr}\right)^\circ\right\}\gamma_0 \\ &= \phi\gamma_j + \zeta\gamma_0, \quad \text{say.} \end{aligned}$$

$$\begin{aligned} Q_0 &= [r_0 - (s_0/k)^+ - (s_0/k)^\circ]\gamma_0 - [(\lambda_0/k)^+ + (\lambda_0/k)^\circ] \sum_{h \neq 0} \gamma_h \\ &= \left\{r_0 - \left(s_0 - \frac{r_0\lambda_0}{r}\right)^+ - \left(s_0 - \frac{r_0\lambda_0}{r}\right)^\circ\right\}\gamma_0 \\ &= \phi'\gamma_0, \quad \text{say.} \end{aligned}$$

With these new values of ϕ, ϕ' and ζ the same formulae as before may be used for evaluating the parameters, the treatment sum of squares, the standard errors, etc. Where, however, simpler formulae are given for type S in terms of other quantities they will not in general apply to type O : SS.

In working out the analysis of variance it is necessary first to remove from the total sum of squares components due to two blocking systems, not one as for type S, but this causes no difficulty because the two are mutually orthogonal. It then remains to apportion W to treatments and the rest to error.

6. Designs of type O:OS, etc.

Such designs exist and could very well be of use, though they do not appear to have any outstanding advantages. For example, if it were a question of applying four treatments to an already determined system of four rows and six columns, designs of both types O:SO and O:PO are available, thus:

O:SO,						
	O	B	O	A	C	A
	B	C	C	O	A	O
	A	O	B	C	O	B
	C	A	A	B	B	C

O:PO,						
	O	B	C	A	O	B
	B	A	O	C	A	C
	A	C	B	O	C	O
	C	O	A	B	B	A

In the design of type O:SO comparisons involving O have an efficiency factor of 140/153 and the others 35/36. In that of type O:PO (O, A) and (B, C) are first associates; comparisons within these pairs have an efficiency factor of 17/18 and others 34/35. In general efficiency the designs are not very different, and the choice between them would largely depend upon the comparisons to be made.

If a design of types O:SO or O:OS is adopted, the analysis of its data is simple. Because there are two blocking systems two components must be removed from the total sum of squares as is done for type O:SS. However, the treatments are disposed in supplemented balance with regard to only one blocking system and only this need be considered in the evaluation of $\mathbf{Q}$, ϕ, ϕ' and ζ, the formulae for type S being appropriate.

Designs also exist of types O:PS, O:SP, O:TS and O:ST, for example,

O:ST,						
	A	O	C	O	O	B
	O	C	O	B	A	O
	B	C	A	B	A	C
	A	A	C	C	B	B
	C	B	B	A	C	A
	O	B	O	A	C	O

O:SP,				
	O	B	A	O
	A	A	B	C
	B	C	O	O
	C	B	B	A
	O	O	C	A
	A	C	C	B

However, none have been found with any obvious advantages.

7. Designs of type O:OO:SSS

In these designs there was initially an orthogonal three-way classification (type O:OO) to which a fourth classification, treatments, has been added so as to be in supplemented balance in relation to the existing three (type O:OO:SSS). The most likely use arises in adding a further set of treatments to a Latin square as in these examples, where capitals represent the original treatments and small letters the new ones:

Aa	Bb	Co	Da
Bo	Aa	Db	Co
Cb	Do	Ba	Ab
Db	Ca	Ao	Bo

Aa	Bo	Cc	Dd
Bd	Ab	Da	Co
Cb	Dc	Ao	Ba
Do	Cd	Bb	Ac

In such instances the values of k, s, λ, etc., will be the same for each scheme of supplementation, and this will be assumed in what follows. However, generalization would be easy if it were ever called for.

If further treatments are added to a $k \times k$ Latin square, the three blocking systems being represented by a cross, a circle and an asterisk, then

$$\mathbf{Q} = \mathbf{T} - [(\mathbf{nB})^{+} + (\mathbf{nB})^{\circ} + (\mathbf{nB})^{*}]/k + 2\mathbf{r}G/N$$

and

$$\phi = r - 3(s-\lambda)/k,$$

$$\phi' = r_0 - 3s_0/k + 3\lambda_0 r_0/kr,$$

$$\zeta = 3(r_0\lambda - r\lambda_0)/kr.$$

Again all the formulae for type S may be used as long as they are expressed in terms of ϕ, ϕ' and ζ. The simplified formulae will not in general apply to the more complicated types. Also, there are now three components to be extracted from the total sum of squares before the remainder can be apportioned between treatments and error.

Where $(k+1)$ treatments are added to an existing $k \times k$ Latin square there does not appear to be any good alternative to the type of design considered here, but if there are $(k-1)$ treatments it is possible in many cases to use a $(k-2, 1^2)$ orthogonal partition as described by Finney (1945, 1946*a*,*b*), and the question arises which design is to be preferred. For the comparison of the control with another treatment the orthogonal partition is always better; for the comparison of two treatments neither of which is the control the supplemented design will have the advantage. For some transformation sets of the original Latin square, either or both of the designs may be impossible.

8. Duals of supplemented designs

If blocks and treatments are interchanged, the resulting design is said to be the dual of the first. Thus, if the supplemented scheme already used in the design of type O : SO above is taken, i.e.

Block			
I	II	III	IV
O	B	A	C
B	C	O	A
O	C	B	A
A	O	C	B
C	A	O	B
A	O	B	C

and the blocks, I–IV transformed to treatments, A′, B′, C′ and O′, and the treatments, A, B, C and O, to blocks, I′–IV′, the result is

Block			
I′	II′	III′	IV′
A′	A′	A′	A′
A′	B′	B′	A′
B′	C′	B′	B′
C′	C′	C′	B′
O′	O′	O′	C′
O′	O′	O′	C′

which is the original design again, i.e. it is a self-dual.

Usually, of course, the position is more complicated, but a study of duals is worth while. For one thing, if a design has an intelligible dual the adjusted block totals and block parameters are easily derived, and a ready check on the computing obtained by calculating the error sum of squares in a second way independent of the first. For another, duals may themselves be of use as designs, not only the duals of type S but of more complicated types as well; they may appear to be complicated, but the error sum of squares can be obtained by calculating that of the original design, a procedure that will necessarily give the same result, the difference between the two schemes lying only in the names given to the classifications. If now the treatments of the duals (blocks, etc., of the original) are ignored, the resulting design is likely to be quite simple, having only one classification if the basic design is of type S, and a new error sum of squares can be calculated. Its excess over the one previously obtained will represent the amount due to the classification ignored, i.e. treatments. The formula for standard errors may be complicated, but it can always be derived, and this should be done before starting the experiment lest the precision prove unsatisfactory.

In general, a design of type S has unequal replication; when dualized to one of type S′ it will have unequal block size. This is not often called for, but sometimes it is useful to have designs available that permit it; for example, when experimenting with litters of differing size.

Summary

Supplemented balance is described, and some uses indicated. A method of analysis is given, both for designs with two-way classifications and for those with more.

I am much indebted to Dr G. H. Freeman and Mr P. Sprent for their interest and helpful comments.

References

Finney, D. J. (1945). Some orthogonal properties of the 4×4 and 6×6 Latin squares. *Ann. Eugen. Lond.*, **12**, 213–9.

Finney, D. J. (1946*a*). Orthogonal partitions of the 5×5 Latin squares. *Ann. Eugen. Lond.*, **13**, 1–3.

Finney, D. J. (1946*b*). Orthogonal partitions of the 6×6 Latin squares. *Ann. Eugen. Lond.*, **13**, 184–96.

Freeman, G. H. (1957). Some experimental designs of use in changing from one set of treatments to another. Part I. *J. R. Statist. Soc.*, Ser. B, **19**, 154–62.

Hoblyn, T. N., Pearce, S. C. & Freeman, G. H. (1954). Some considerations in the design of successive experiments in fruit plantations. *Biometrics*, **10**, 503–15.

Pearce, S. C. (1952). Some new designs of Latin square type. *J. R. Statist. Soc.*, Ser. B, **14**, 101–6.

Pearce, S. C. (1953). Field experimentation with fruit trees and other perennial plants. *Tech. Commun. Commonwealth Bur. Hort. Plantation Crops*, **23**.

Tocher, K. D. (1952). The design and analysis of block experiments. *J. R. Statist. Soc.*, Ser. B, **14**, 45–100.

Yates, F. (1936). Incomplete Latin squares. *J. Agric. Sci.* **26**, 301–15.

Biometrika (1960), **47**, 3 *and* 4, *p.* 273
Printed in Great Britain

The efficiency of blocking in incomplete block designs*

BY JOHN LEROY FOLKS
Texas Instruments Incorporated and Statistical Laboratory, Iowa State College
AND OSCAR KEMPTHORNE
Statistical Laboratory, Iowa State College

1. *Summary.* Procedures for estimating the relative efficiency of certain designs have been considered by Yates (1935), Cochran & Cox (1957), and Kempthorne (1952, 1955). In this paper the efficiency of blocking in general is considered. A general formula for any incomplete block design of fairly general form is obtained.

2. *Introduction.* In many experiments one has the option of choosing from a wide class of designs ranging from designs involving no blocking to designs involving two-way classification of experimental units. A measure of the efficiency of one design relative to another is the inverse ratio of the variances with which treatment differences are estimated in the two designs. Given that design A has been used, estimation of its efficiency relative to some other design, B say, requires that we estimate the error variance of design B from the data obtained from design A.

A common elementary situation is that a randomized block design has been used and it is desired to estimate the error variance had a completely randomized design been used. The formula for this purpose is well known, but we wish to summarize a few of the salient points of the randomization theory upon which it is based. Throughout we shall concern ourselves only with variability over which we randomize.

Consider the conceptual population of yields that would be obtained if every treatment could be applied to every plot. Denote the yield with treatment k $(= 1, 2, \ldots, t)$ on plot j $(= 1, 2, \ldots, t)$ of block i $(= 1, 2, \ldots, r)$ by X_{ijk}. Then we may write an identity

$$X_{ijk} = X_{...} + (X_{i..} - X_{...}) + (X_{ijk} - X_{ij.}) + (X_{ij.} - X_{i..}),$$

where the dots indicate averages over the subscripts omitted. The estimation of variance under randomization of observed mean treatment differences is possible only if additivity of treatment and plot effects holds, or, with another mode of expression, the treatment effects do not depend on the plot. The additivity assumption leads to

$$X_{ijk} - X_{ij.} = t_k \quad \text{for all } i, j.$$

In order to develop the argument on efficiency of blocking this relationship must be assumed. We can then write

$$X_{ijk} = \mu + b_i + t_k + e_{ij}$$

with an obvious identification of symbols. However, we cannot observe the yield of treatment k on every plot but only the yield of treatment k upon a randomly chosen plot in each of the blocks. For such considerations the design random variables (Kempthorne 1952, 1955) described below have been found helpful. Let

$\delta_{ij}^k = 1$ if treatment k is assigned to plot j in the ith block
$= 0$ otherwise.

* Journal Paper Number J-3571 of the Iowa Agricultural and Home Economics Experiment Station, Ames. Project 890.

From elementary probability considerations we have

$$E(\delta_{ij}^k) = 1/t,$$
$$\delta_{ij}^k \delta_{ij'}^k = 0,$$
$$\delta_{ij}^k \delta_{ij'}^{k'} = 0,$$
$$E(\delta_{ij}^k \delta_{ij'}^{k'}) = 1/t(t-1), \quad \text{etc.}$$

Denote the observed value of treatment k in block i by y_{ik}. Then

$$y_{ik} = \sum_j \delta_{ij}^k X_{ijk}$$
$$= \mu + b_i + t_k + \sum_j \delta_{ij}^k e_{ij}.$$

The random variables in this expression are the δ_{ij}^k, and their distribution is determined by the randomization procedure. This allows us to evaluate the expected sums of squares appearing in the ordinary analysis of variance. For instance, the treatment sum of squares is equal to

$$\sum_{ik} (y_{.k} - y_{..})^2 = \sum_{ik} \left(t_k + \frac{1}{r} \sum_{ij} \delta_{ij}^k e_{ij} \right)^2$$

and it can be verified that

$$E \text{ (treatment } SS) = r \sum_k t_k^2 + E \frac{1}{r^2} \sum_k \left(\sum_{ij} \delta_{ij}^k e_{ij} \right)^2$$
$$= r \sum_k t_k^2 + \frac{1}{r} \sum_{ij} e_{ij}^2.$$

The expected values of sums of squares for the analysis of variance of randomized blocks are shown in Table 1, and in particular the expected error variance is

$$\frac{1}{r(t-1)} \sum_{ij} e_{ij}^2.$$

Table 1. *Analysis of variance of randomized block design*

Due to ...	d.f.	$E(SS)$	m.s.
Blocks	$r-1$	$t \sum_i b_i^2$	B
Treatments	$t-1$	$\frac{1}{r} \sum_{ij} e_{ij}^2 + r \sum_k t_k^2$	—
Error	$(r-1)(t-1)$	$\frac{r-1}{r} \sum_{ij} e_{ij}^2$	E
Total	$rt-1$	$\sum_{ij} e_{ij}^2 + t \sum_i b_i^2 + r \sum_k t_k^2$	—

If a completely randomized design had been used we would be observing X_{ijk} for r combinations of i and j for each value of k without the restriction of randomized blocks that we get only one ij combination with a particular value of i for each k. So if we again use design random variables δ_{ij}^k the property

$$\delta_{ij}^k \delta_{ij'}^k = 0$$

no longer holds. In the randomized blocks design treatment k cannot occur on both of plots j and j' of block i. But when we randomize ignoring the blocking, plots in the same or

in different blocks are given exactly the same status. The mechanics of working out the consequences of the different properties of the δ's need not be presented here (Kempthorne, 1955). They lead to the analysis of variance in Table 2. The expected error variance in this case is

$$(\sum_{ij} e_{ij}^2 + t \sum_i b_i^2)/(rt-1).$$

From Table 1 we see that this quantity can be estimated by

$$\frac{(r-1)B + r(t-1)E}{rt-1}.$$

Table 2. *Analysis of variance of completely randomized design*

Due to ...	d.f.	$E(SS)$
Treatments	$t-1$	$\frac{t-1}{rt-1}(\sum_{ij} e_{ij}^2 + t\sum_i b_i^2) + r\sum_k t_k^2$
Error	$t(r-1)$	$\frac{t(r-1)}{rt-1}(\sum_{ij} e_{ij}^2 + t\sum_i b_i^2)$
Total	$rt-1$	$\sum_{ij} e_{ij}^2 + t\sum_i b_i^2 + r\sum_k t_k^2$

The notable fact about this expression is that it is obtainable by simulating an experiment in which the treatments have no effect by using an actual experiment in which there are treatment effects. In the case of a completely randomized design using the same plots, the expected error mean square is equal to the total mean square which would be observed with dummy treatments. This is therefore estimated by pooling the sum of squares for blocks, the estimated sum of squares for treatments under the assumption of no treatment effects, and the sum of squares for error, i.e.

$$\frac{(r-1)B + (t-1)E + (r-1)(t-1)E}{rt-1}.$$

For this type of argument, which was used first by Yates (1935), to be valid it is necessary that the designs under consideration be unbiased designs, i.e. designs for which the expected value of the treatment mean square with zero treatment effects is equal to the expected value of the error mean square. It is important to realize that the substitution of error mean square for treatment mean square and pooling of resulting sums of squares is merely a computational algorithm, the justification of which can be obtained by the expectations of mean squares under randomization.

The problem we are concerned with is the efficiency of the blocking in any incomplete block design. We suppose that the results of an incomplete block design are available, and we wish to estimate what the error variance would have been had we not used the blocking. This information alone would not tell us the actual efficiency of the incomplete block experiment relative to an unrestricted experiment. In the incomplete block design there may, and usually will, be treatment comparisons of various precisions, these precisions depending on the particular way the confounding with blocks affects each possible treatment comparison. The loss of information due to confounding alone is conveyed by the efficiency factor of the design or by the efficiency factor for each possible comparison. These efficiency factors are consequences solely of the structure of the incomplete block design, are usually specified when the method of analysis of an incomplete block design is presented,

and are entirely unrelated to the efficacy of the blocking in controlling variability in a particular experimental situation. For example, to estimate the mean variance of intra-block estimates of treatment differences for the incomplete block design in which each treatment is represented r times we use the formula

$$2(\text{intra block error mean square})/(rE),$$

where E is the overall efficiency factor of the design. The mean variance of estimated treatment differences if no blocking had been used would equal

$$2(\text{unrestricted error mean square})/r.$$

We will know the intra-block error mean square for the incomplete block design and the efficiency factor of the design and to estimate the loss or gain from the blocking we need to estimate the unrestricted error mean square, i.e. the error mean square if no blocking had been used.

3. *Randomization analysis of incomplete block designs.* We shall consider an arbitrary incomplete block design, the only restrictions being that each treatment is represented r times and the block size is a constant k. This includes the case of balanced incomplete block designs, partially balanced incomplete block designs (p.b.i.b.), and rectangular lattices which are not p.b.i.b. designs. It is convenient to use a formulation in which one distinguishes a design and a master plan. A master plan is a specification of the arrangement of treatments in sets or arrays. A design is generated from a master plan by using a randomization procedure which specifies the probabilities of the possible combinations of the treatments in the plan and the pieces of experimental material. The master plan plus the randomization instructions comprise the design. A particular realization of the process is the plan of the experiment in the ordinary sense of the word.

The situation is summarized as follows:

(*a*) There are r groups ($i = 1, 2, \ldots, r$) of blocks ($j = 1, 2, \ldots, b$) and each block has k plots ($m = 1, 2, \ldots, k$).

(*b*) The master plan is of the form

(1) There are r subplans ($i' = 1, 2, \ldots, r$).

(2) In each subplan there are b sets ($u = 1, 2, \ldots, b$) of k of the treatments ($r = 1, 2, \ldots, k$). A treatment is not necessarily confined to one set, though for resolvable plans this is the case. There is no necessity for the master plan to be balanced except that each treatment is to occur R times in each subplan.

(*c*) The randomization is:

(1) Subplans are assigned to groups arbitrarily.

(2) Given a group of blocks and the subplan assigned to it, sets of treatments are assigned to blocks at random.

(3) Given a block and the set of treatments assigned to it, the treatments in the set are assigned at random to the plots of the blocks.

All analyses that we consider are made within groups of blocks so we shall ignore the first part of the randomization and identify subplans and groups by the same subscript i. The population model of conceptual yields assuming additivity (of treatments and plots) is

$$X_{ijmw} = \mu + r_i + b_{ij} + e_{ijm} + t_{iw},$$

in which $i = 1, 2, \ldots, r$ denotes group and subplan;
$j = 1, 2, \ldots, b$ denotes block within group;
$m = 1, 2, \ldots, k$ denotes plot within block;
$u = 1, 2, \ldots, b$ denotes set within subplan;
$v = 1, 2, \ldots, k$ denotes treatment within set.

The following relationships hold among the quantities in the model

$$\sum_i r_i = \sum_j b_{ij} = \sum_m e_{ijm} = \sum_{uv} t_{iuv} = 0.$$

Let $\alpha_{ij}^{iu} = 1$ if the uth set is assigned to the jth block in the ith group
$= 0$ otherwise,

$\beta_{ijm}^{iuv} = 1$ if the uvth treatment is assigned to plot m of the jth block of the ith group
$= 0$ otherwise.

The observed yields are given by

$$\begin{aligned} y_{iuv} &= \sum_{jm} \alpha_{ij}^{iu} \beta_{ijm}^{iuv} X_{ijmuv} \\ &= \mu + r_i + t_{iuv} + \sum_j \alpha_{ij}^{iu} b_{ij} + \sum_{jm} \alpha_{ij}^{iu} \beta_{ijm}^{iuv} e_{ijm} \\ &= \mu + r_i + t_{iuv} + \beta_{iu} + \gamma_{iuv}, \quad \text{say.} \end{aligned}$$

It can be shown that

$$E(\beta_{iu}) = 0, \quad E(\beta_{iu}^2) = \frac{1}{b} \sum_j b_{ij}^2, \quad E(\sum_i \beta_{iu})^2 = \frac{1}{b} \sum_{ij} b_{ij}^2,$$

$$E(\sum_i \beta_{iu})(\sum_i \beta_{iu'}) = \frac{1}{b(b-1)} \sum_{ij} b_{ij}^2 \quad (u \neq u')$$

and that

$$E(\gamma_{iuv}) = 0, \quad E(\gamma_{iuv}^2) = \frac{1}{bk} \sum_{jm} e_{ijm}^2,$$

$$E(\sum_i \gamma_{iuv})^2 = \frac{1}{bk} \sum_{ijm} e_{ijm}^2$$

$$E(\sum_i \gamma_{iuv})(\sum_i \gamma_{iu'v'}) = 0 \quad (u \neq u').$$

We shall now find expected values of sums of squares appearing in the analyses of variance.

$$\begin{aligned} E\,(\text{total } SS) &= E \sum_{iuv} (y_{iuv} - y_{...})^2 \\ &= E \sum_{iuv} (r_i + t_{iuv} + \beta_{iu} + \gamma_{iuv})^2 \\ &= bk \Sigma r_i^2 + \sum_{iuv} t_{iuv}^2 + k \sum_{ij} b_{ij}^2 + \sum_{ijm} e_{ijm}^2, \\ E\,(\text{group } SS) &= bk \sum_i r_i^2, \\ E\,(\text{block ignoring treatment } SS) &= E \sum_{iuv} (y_{iu.} - y_{i..})^2 \\ &= E \sum_{iuv} \left(\frac{1}{k} \sum_v t_{iuv} + \beta_{iu} + \frac{1}{k} \sum_v \gamma_{iuv} \right)^2 \\ &= \frac{1}{k} \sum_{iu} (\sum_v t_{iuv})^2 + k \sum_{ij} b_{ij}^2. \end{aligned}$$

The expected error sum of squares can be obtained by using the fact that the incomplete block designs considered are unbiased. This gives

$$E(\text{error } SS) = \frac{rbk - rb - t + 1}{rb(k-1)} \sum_{ijm} e_{ijm}^2.$$

The expectation of the adjusted treatment sum of squares can be found by subtraction to give analysis of variance I.

Table 3. *Analysis of variance I of incomplete block design*

Due to ...	d.f.	$E(SS)$
Groups	$r-1$	$bk \sum_i r_i^2$
Blocks ignoring treatments	$r(b-1)$	$\frac{1}{k}\sum_{iu}(\sum_v t_{iuv})^2 + k\sum_{ij} b_{ij}^2$
Treatments adjusted for blocks	$t-1$	$\frac{t-1}{rb(k-1)}\sum_{ijm} e_{ijm}^2 + \sum_{iuv} t_{iuv}^2 - \frac{1}{k}\sum_{iu}(\sum_v t_{iuv})^2$
Error	$rbk-rb-t+1$	$\frac{rbk-rb-t+1}{rb(k-1)}\sum_{ijm} e_{ijm}^2$
Total	$rbk-1$	—

To obtain the expectation of the unadjusted treatment SS, let

$\delta_{iuv}^w = 1$ if the wth treatment corresponds with iuv index
$= 0$ otherwise,

$\delta_{iu} = 1$ if the wth treatment occurs in the uth block in the ith group
$= 0$ otherwise,

where

$$\sum_v \delta_{iuv}^w = \delta_{iu}^w$$

$$\sum_u \delta_{iu}^w = R,$$

E(unadjusted treatments SS)

$$= E\frac{1}{rR}\sum_w \left(\sum_{iuv} \delta_{iuv}^w y_{iuv} - \frac{1}{t}\sum_w \sum_{iuv} \delta_{iuv}^w y_{iuv}\right)^2$$

$$= \sum_{iuv} t_{iuv}^2 + \frac{1}{rR} E \sum_w (\sum_{iu} \delta_{iu}^w \beta_{iu})^2 + \frac{1}{rR} E \sum_w (\sum_{iuv} \delta_{iuv}^w \gamma_{iuv})^2$$

$$= \sum_{iuv} t_{iuv}^2 + \frac{t(b-R)}{rb(b-1)} \sum_{ij} b_{ij}^2 + \frac{t}{rbk} \sum_{ijm} e_{ijm}^2.$$

This enables us to compute the expectations for analysis of variance II, which is required to estimate the error variance had no blocking been performed. Table 3 is useful in that the sum of squares for blocks ignoring treatments plus the sum of squares for treatments adjusted for blocks is equal to the sum of squares for treatments ignoring blocks plus the sum of squares for blocks eliminating treatments. The easiest way to get Table 4 is by way of Table 3 and the expectation of the sum of squares for treatments ignoring blocks.

Had no blocking been done within groups, the analysis of variance with dummy treatments would have been that given in Table 5. The proper error variance for this case is

$$\frac{1}{r(bk-1)}\left(\sum_{ijm} e_{ijm}^2 + k\sum_{ij} b_{ij}^2\right).$$

This is estimated by

$$E' = \frac{bk}{bk-1}\,\frac{k-1}{k}\,E + \frac{r(b-1)B - \left(\frac{t-k}{k}\right)E}{rbk - \frac{t(b-R)}{b-1}}.$$

Table 4. *Analysis of variance II for incomplete block design*

Due to ...	d.f.		m.s.
Groups	$r-1$	$bk\sum_i r_i^2$	—
Blocks eliminating treatments	$r(b-1)$	$\frac{t-k}{rbk(k-1)}\sum_{ijm} e_{ijm}^2$	—
		$+\frac{rbk(b-1)-t(b-R)}{rb(b-1)}\sum_{ij} b_{ij}^2$	B
Treatments ignoring blocks	$t-1$	$\frac{t}{rbk}\sum_{ijm} e_{ijm}^2 + \frac{t}{r}\frac{(b-R)}{b(b-1)}\sum_{ij} b_{ij}^2$	—
		$+\sum_{iuv} t_{iuv}^2$	
Error	$rbk-rb-t+1$	$\frac{rbk-rb-t+1}{rb(k-1)}\sum_{ijm} e_{ijm}^2$	E
Total	$rbk-1$	—	—

Table 5. *Analysis of variance with dummy treatments for no blocking within groups*

Due to ...	d.f.	$E(SS)$
Groups	$r-1$	$bk\sum_i r_i^2$
Within groups	$r(bk-1)$	$\sum_{ijm} e_{ijm}^2 + k\sum_{ij} b_{ij}^2$
Total	$rbk-1$	$\sum_{ijm} e_{ijm}^2 + k\sum_{ij} b_{ij}^2 + bk\sum_i r_i^2$

Special cases

(1) When $R = 1$, the formula reduces to a simple pooling of the sums of squares for blocks eliminating treatments and for error. This includes all the lattice designs, rectangular lattices, and confounded factorials arranged in complete replicates.

(2) Randomized block with r blocks, t treatments. The formula given above reduces to

$$E' = \frac{rt}{rt-1}\left\{\frac{t-1}{t}E + \frac{(r-1)B}{rt}\right\} = \frac{r(t-1)E + (r-1)B}{rt-1}$$

which agrees with the formula in § 2.

Example $r = 1,\quad t = 13,\quad b = 13,\quad R = 4,\quad k = 4,$

$$E' = \frac{52}{51}\left\{3E/4 + \frac{12B - 9E/4}{52 - 39/4}\right\} = (64B + 157E)/(13)(17).$$

With the incomplete block design the variance of treatment differences is estimated to be

$$2E\Big/\frac{R(k-1)t}{k(t-1)} = 8E/13.$$

The variance of treatment differences had a completely randomized design been used is estimated to be

$$2E'/R = 2E'/4.$$

Hence the relative efficiency of the incomplete block experiment is estimated to be

$$\frac{E'}{4}\Big/\frac{4E}{13} = 13E'/(16E) = \frac{4B}{17E} + \frac{157}{(16)(17)}.$$

This estimate of relative efficiency does not utilize the interblock information and is therefore unreliable if interblock information is usable.

4. *Randomization analysis of designs involving two-way classification of experimental units.* The experimental situation is summarized as follows:

(*a*) There are r groups ($i = 1, 2, \ldots, r$) of bk plots, and each group consists of k rows ($j = 1, 2, \ldots, k$) and b ($m = 1, 2, \ldots, b$) columns.

(*b*) The master plan consists of r sets within each of which each of t treatments ($w = 1, 2, \ldots, t$) is represented R times. Within each set the treatments are arranged in a $b \times k$ array and a treatment in the ith group is identified by u, its row number, and v, its column number.

(*c*) The randomization is

(1) Sets are assigned to groups arbitrarily, and the same index i is used to designate group of plots and set of treatments.

(2) Within each group the rows and columns of the treatment array are assigned at random to the rows and columns of the plot array.

The population model of conceptual yields is

$$X_{ijmuv} = \mu + s_i + r_{ij} + c_{im} + t_{iuv} + e_{ijm},$$

with

$$\sum_i s_i = \sum_j r_{ij} = \sum_m c_{im} = \sum_j e_{ijm} = \sum_m e_{ijm} = \sum_{uv} t_{iuv} = 0.$$

Let $\alpha_{ij}^{iu} = 1$ if uth row of the ith treatment array is assigned to the mth column of experimental units in the ith group

$= 0$ otherwise,

$\beta_{im}^{iv} = 1$ if vth column of the ith treatment array is assigned to the mth column of experimental units in the ith group

$= 0$ otherwise.

The observed yields are given by

$$y_{iuv} = \sum_{jm} \alpha_{ij}^{u} \beta_{im}^{v} X_{ijmuv} = \mu + s_i + \alpha_{iu} + \beta_{iv} + \gamma_{iuv} + t_{iuv}.$$

We can easily obtain the expectations of certain sums of squares.

$$E(\text{total } SS) = bk \sum_i s_i^2 + b \sum_{ij} r_{ij}^2 + k \sum_{im} c_{im}^2 + \sum_{ijm} e_{ijm}^2 + rR \sum_w t_w^2,$$

$$E(\text{group } SS) = bk \sum_i s_i^2.$$

Let $\delta_{iuv}^{w} = 1$ if wth treatment = iuvth treatment

$= 0$ if otherwise,

E (unadjusted treatment SS)

$$= \frac{1}{rR} E \sum_{w} [\sum_{iuv} \delta_{iuv}^{w} y_{iuv} - (1/t) \sum_{w} \sum_{iuv} \delta_{iuv}^{w} y_{iuv}]^2$$

$$= rR \sum_{w} t_w^2 + \frac{t}{r} \frac{k-R}{k(k-1)} \sum_{ij} r_{ij}^2 + \frac{t}{r} \frac{b-R}{b(b-1)} \sum_{im} c_{im}^2$$

$$+ \frac{t}{rbk} \frac{(k-1)(b-1)+(R-1)}{(k-1)(b-1)} \sum_{ijm} e_{ijm}^2.$$

We then have the analysis of variance in Table 6.

Table 6. *Analysis of variance for two-way classification*

Due to ...	d.f.	$E(SS)$
Groups	$r-1$	$bk \sum_i s_i^2$
Rows and columns eliminating treatments	$r(b+k-2)$	$\frac{t(b+k-R)-bk}{rbk(k-1)(b-1)} \sum_{ijm} e_{ijm}^2 + \frac{rbk-t}{rk} \frac{k-R}{k-1} \sum_{ij} r_{ij}^2 + \frac{rbk-t}{rb} \frac{b-R}{b-1} \sum_{im} c_{im}^2$
Treatments ignoring rows and columns	$t-1$	$\frac{t}{rbk}\left[1+\frac{R-1}{(b-1)(k-1)}\right] \sum_{ijm} e_{ijm}^2 + \frac{t}{rk} \frac{k-R}{k-1} \sum_{ij} r_{ij}^2 + \frac{t}{rb} \frac{b-R}{b-1} \sum_{im} c_{im}^2 + rR \sum_{w} t_w^2$
Error	$r(b-1)(k-1)-(t-1)$	$\frac{r(k-1)(b-1)-(t-1)}{r(b-1)(k-1)} \sum_{ijm} e_{ijm}^2$
Total	$rbk-1$	$\sum_{ijm} e_{ijm}^2 + b \sum_{ij} r_{ij}^2 + k \sum_{im} c_{im}^2 + bk \sum_i s_i^2 + rR \sum_w t_w^2$

If treatments were completely randomized within groups the error variance would be

$$\frac{1}{r(bk-1)} (\sum_{ijm} e_{ijm}^2 + k \sum_{im} c_{im}^2 + b \sum_{ij} r_{ij}^2).$$

For $R = 1$, this is estimated by pooling the error and rows and columns (eliminating treatments) sums of squares. This is the case for lattice squares.

For the Youden squares, including Latin squares and extended Youden squares, blocking may be done by rows or columns alone, as opposed to complete randomization.

Consider the Youden squares where every treatment occurs in every row, i.e.

$$b = t, \quad R = k.$$

We have the analysis of variance given in Table 7.

Table 7. *Analysis of variance for Youden square*

Due to ...	d.f.	$E(SS)$	m.s.
Groups	$r-1$	$bk\sum_{ij} s_i^2$	A
Rows	$r(k-1)$	$b\sum_{ij} r_{ij}^2$	—
Columns eliminating treatments	$r(b-1)$	$\frac{b-k}{rk(k-1)(b-1)}\sum_{ijm} e_{ijm}^2 + \left[k-\frac{b-k}{r(b-1)}\right]\sum_{im} c_{im}^2$	B
Treatments ignoring rows and columns	$t-1$	$\frac{b}{rk(b-1)}\sum_{ijm} e_{ijm}^2 + \frac{b-k}{r(b-1)}\sum_{im} c_{im}^2 + rR\sum_{w} t_w^2$	—
Error	$[r(k-1)-1][b-1]$	$\left[1-\frac{1}{r(k-1)}\right]\sum_{ijm} e_{ijm}^2$	E
Total	$rbk-1$	$bk\sum_{i} s_i^2 + b\sum_{ij} r_{ij}^2 + k\sum_{im} c_{im}^2 + \sum_{ijm} e_{ijm}^2 + rR\sum_{w} t_w^2$	—

In this case the error variance if treatments had been completely randomized within groups is estimated by

$$E' = \frac{1}{r(bk-1)}\left\{r(b-1)(k-1)E + r(k-1)A + \frac{r(b-1)B - \left(\frac{b-k}{k}\right)E}{\frac{rk(b-1)-(b-k)}{rk(b-1)}}\right\}.$$

The error variance for randomized blocks within groups with blocks made up by rows is estimated by

$$E' = \frac{1}{r(bk-1)-(k-1)}\left\{r(b-1)(k-1)E + \frac{r(b-1)B - \left(\frac{b-k}{k}\right)E}{\frac{rk(b-1)-(b-k)}{rk(b-1)}}\right\}.$$

If blocking is done by columns, the error variance is estimated by

$$E' = \frac{1}{r(bk-1)-(b-1)}\{r(k-1)A + r(k-1)(b-1)E\}.$$

It should be noted that when $k = b = t, \quad r = 1,$

the preceding formulae reduce to the usual ones given for the Latin squares.

As an example consider a Youden square for which

$$R = 4, \quad r = 1, \quad b = t = 7, \quad \text{and} \quad k = 4;$$

A = row mean square;
B = column eliminating treatments mean square;
E = error mean square.

Had a completely randomized design been used

$$E' = \frac{1}{27}\left\{18E + 3A + \frac{6B - 3E/4}{(24-3)/24}\right\} = \tfrac{1}{9}(40E/7 + A + 16B/7).$$

Had blocking been done by rows

$$E' = \frac{1}{24}\left\{18E + \frac{6B - 3E/4}{\frac{21}{24}}\right\} = \tfrac{1}{8}(40E/7 + 16B/7).$$

Had blocking been done by columns

$$E' = \tfrac{1}{21}[3A + 18E] = \tfrac{1}{7}[A + 6E].$$

References

Cochran, W. G. & Cox, Gertrude M. (1957). *Experimental Designs.* New York: John Wiley and Sons.

Kempthorne, O. (1952). *The Design and Analysis of Experiments.* New York: John Wiley and Sons.

Kempthorne, O. (1955). The randomization theory of experimental inference. *J. Amer. Statist. Ass.* **50**, 946–67.

Yates, F. (1935). Complex experiments. *J.R. Statist. Soc.* Suppl. **2**, 181–247.

Biometrika (1960), **47**, 3 *and* 4, *p*, 285
Printed in Great Britain

Queueing with balking. II

BY FRANK A. HAIGHT
Institute of Transportation and Traffic Engineering, University of California, Los Angeles

1. INTRODUCTION

We consider a stochastic model proposed in an earlier paper (Haight, 1957, which will be referred to as QWB) which may be described briefly as follows: arrivals to a queue occur at instants which constitute a homogeneous Poisson process with parameter λ. Each arrival will join the queue if its length is K or less and otherwise balk, i.e. go away and not return. The distribution of K is denoted by $F(n) = \Pr(K \leqslant n)$, $G(n) = 1 - F(n)$. The distribution of service times is $1 - e^{-\mu x}$, and we are mainly interested in equilibrium.

The decisions (to join or balk) form a sequence similar to those studied in the theory of runs, but without the restriction that the results arise from independent Bernoulli trials.

It is not intuitively clear whether or not a long sequence of balks should predispose a new arrival to balk. On one hand such a sequence argues that the queue length has been reduced to some small value; on the other hand it should have been large to produce such a sequence. Furthermore, the reduction in queue length during a sequence of balks is produced by the service mechanism, and so a knowledge of the join-balk sequence is apparently insufficient to determine whether a new arrival will balk, since that decision is postulated to depend solely on queue length.

The device which makes such an investigation possible is the use of the waiting-time distribution. Waiting time decreases equally with elapsed time, except when an arrival joins the queue. It is thus independent of the instants of service termination, and depends only on the join-balk sequence. Furthermore, its distribution can be expressed in terms of the queue length distribution and the constants λ and μ.

Let the system be said to be in the state E_n, $n = 0, 1, 2, \ldots$ if at time t there are n customers present, including the one if any then being served. Let $\tau_1, \tau_2, \ldots, \tau_n, \ldots$ denote the points at which customers arrive, so that $\{\tau_n\}$ is a Poisson process. Define random variables $k(t)$ as follows: $k(t) = n$ if at time t the system is in state E_n. Define a random variable $w(t)$ as follows

$$w(t) = s_1 + s_2 + \ldots + s_{k(t)},$$

where $s_1, s_2, \ldots$ are independently and identically distributed random variables with common distribution function $1 - e^{-\mu x}$, for $x \geqslant 0$.† $w(t)$ is called the waiting time at time t. Define random variables k_n, w_n as follows: $k_n = k(\tau_n - 0)$, $w_n = w(\tau_n - 0)$. Thus k_n is the queue length just before the nth Poisson point τ_n and w_n is the waiting time just before that point.

Place $$P(n, t) = \Pr(k(t) = n), \quad Q^m(n) = \Pr(k_m = n).$$

† If $k(t) \geqslant 1$, $s_{k(t)}$ is the residual service time of the customer being served. If $k(t) > 1$, $s_{k(t)-1}, \ldots, s_1$ are the service times of the waiting customers.

Then it can be shown that $p(n) = \lim_{m\to\infty} Q^m(n) = \lim_{t\to\infty} P(n,t)$ exists under the conditions of QWB. Similarly, place

$$H(t,x) = \Pr(w(t) \leqslant x), \quad H^m(x) = \Pr(w_m \leqslant x).$$

Then it can be shown that $H(x) = \lim_{m\to\infty} H^m(x) = \lim_{t\to\infty} H(t,x)$ exists under the conditions of QWB. Following the notation of QWB further, we define

$$\eta(s) = \sum_0^\infty p(n)\, s^n, \quad \xi(s) = \sum_0^\infty p(n+1)\, F(n)\, s^n.$$

The density function $h(x)$ which corresponds to $H(x)$ has the value $p(0)$ at the origin, since there is no waiting time when the queue is empty. In the opposite case, the waiting time is the sum of n negative exponential phases, where n is the number waiting or being served on arrival. Therefore

$$h(x) = p(0)\, j(x) + \sum_1^\infty \frac{\mu^n x^{n-1} e^{-\mu x}}{\Gamma(n)} p(n), \tag{1}$$

where $j(x)$ is the Dirac delta function. If we write the incomplete gamma function

$$\gamma(n,y) = \int_0^y e^{-t} t^{n-1} dt = \Gamma(n) - \Gamma(n,y),$$

then it is easy to see that

$$H(x) = p(0) + \sum_1^\infty \frac{\gamma(n,\mu x)}{\Gamma(n)} p(n). \tag{2}$$

It will also be useful to define a function similar to (2), but with the queue index advanced one unit

$$J(x) = p(1) + \sum_1^\infty \frac{\gamma(n,\mu x)}{\Gamma(n)} p(n+1). \tag{3}$$

Using an asterisk to denote Laplace transformations, we have

$$h^*(s) = p(0) + \sum_1^\infty p(n) \left[\frac{\mu}{s+\mu}\right]^n$$

$$= \eta\left[\frac{\mu}{s+\mu}\right] \tag{4}$$

as is well known (see, for example, Gaver, 1954). Also

$$H^*(s) = h^*(s)/s, \tag{5}$$

omitting the term $h(0+)$ because of the presence of $j(x)$, as mentioned, for example, by Truxal (1955). The transformation of $J(x)$ may be computed similarly, using the formula

$$p(n+1) = \rho G(n-1)\, p(n) \quad (\rho = \lambda/\mu)$$

from QWB; the result is

$$J^*(s) = \frac{\mu+s}{\mu s}[h^*(s) - p(0)]. \tag{6}$$

2. Conditional distributions

We now consider the queue length and waiting-time distributions, together with the generating functions and Laplace transformations associated with them, under certain specific conditions. These conditions will be indicated by subscripts, according to the following system:

B as a subscript indicates the situation at the time of arrival of a customer who balks.

J as a subscript indicates the situation at the time of arrival of a customer who joins.

$N = 1, 2, \ldots$, indicates the situation conditional on the arrival of a customer who is preceded by exactly $N-1$ persons balking. That is, the time between this customer's arrival and the preceding addition to the queue consists of N negative exponential phases. The subscript does not indicate whether the arrival himself balks or joins.

Each of these functions has its corresponding $p(n,t)$, $Q^m(n)$, $H(t,x)$ and $H^m(x)$, as well as limiting theorems analogous to those quoted in the preceding section. In all cases, we are interested only in the limiting (equilibrium) expressions and therefore pass directly from probabilistic definitions to equilibrium equations.

It is shown in QWB that $\xi(1)$ is the probability that an individual balks; this quantity will be abbreviated ξ. Then

$$\begin{aligned} H_J^m(x) &= \Pr[w_m \leqslant x | k(\tau_m+0) = k(\tau_m-0)+1] \\ &= \frac{1}{1-\xi}\Pr(w_m \leqslant x)\Pr[k(\tau_m+0) = k(\tau_m-0)+1 | w_m \leqslant x]. \end{aligned} \tag{7}$$

The last factor may be evaluated as follows

$$\begin{aligned} &\Pr[k(\tau_m+0) = k(\tau_m-0)+1 | w_m \leqslant x] \\ &\quad = \sum_{n=0}^{\infty} \Pr[k(\tau_m+0)+k(\tau_m-0)+1 | k(\tau_m-0) = n]\Pr[k(\tau_m-0) = n | w_m \leqslant x] \\ &\quad = \sum_{n=0}^{\infty} G(n-1)\Pr[k(\tau_m-0) = n | w_m \leqslant x]. \end{aligned}$$

In the limiting case this becomes

$$\begin{aligned} &= \frac{1}{H(x)}\sum_{1}^{\infty} G(n-1)\,p(n)\frac{\gamma(n,\mu x)}{\Gamma(n)} + \frac{p(0)}{H(x)} \\ &= \frac{J(x)}{\rho H(x)}. \end{aligned}$$

Substituting this into (7), we have

$$H_J(x) = \frac{J(x)}{\rho(1-\xi)}. \tag{8}$$

Using (5) with the J subscript, and (6), we obtain

$$h_J^*(s) = \frac{\mu+s}{\lambda(1-\xi)}[h^*(s) - p(0)]. \tag{9}$$

(4) is equivalent to

$$\eta_J(s) = h_J^*\left[\frac{\mu(1-s)}{s}\right] \tag{10}$$

and with (9) this gives

$$\eta_J(s) = \frac{\eta(s) - p(0)}{\rho s(1-\xi)}. \tag{11}$$

Expanding into a series in s, we find that

$$p_J(n) = \frac{p(n+1)}{\rho(1-\xi)} \quad (n = 0, 1, 2, \ldots). \tag{12}$$

Denoting the mean by m [and using the formula $1-p(0) = \rho(1-\xi)$ from QWB†] we have

$$m_J = \frac{m}{\rho(1-\xi)} - 1. \tag{13}$$

It is easy to see that m_J is always positive. For $\rho(1-\xi)$ is the effective traffic intensity ρ' (as explained in QWB), and therefore

$$m = \frac{\rho'}{1-\rho'} > \rho'.$$

Using (11) higher moments of the conditional distribution can be obtained.

The various quantities associated with the conditional distribution of individuals who balk can be obtained by arguments like those used above, or else by noting that, for any relevant function f,

$$\xi f_J + (1-\xi) f_B = f. \tag{14}$$

Some results are

$$H_B(x) = \frac{H(x)}{\rho} - \frac{J(x)}{\rho\xi}, \tag{15}$$

$$\eta_B(s) = \frac{\eta(s)}{\xi} - \frac{\eta(s)}{\rho\xi s} + \frac{p(0)}{\rho\xi s}, \tag{16}$$

$$\left.\begin{aligned} p_B(0) &= 0, \\ p_B(n) &= \frac{1}{\xi}\left[p(n) - \frac{1}{\rho}p(n+1)\right] \quad (n > 0), \end{aligned}\right\} \tag{17}$$

and

$$m_B = \frac{m}{\xi} - \frac{m-1}{\rho\xi} - \frac{p(0)}{\rho\xi}. \tag{18}$$

Next we consider the situation in which an arrival is preceded by $N-1$ balks, saying nothing about whether or not this arrival himself balks, and find the probability distribution of queue length. First, suppose $N = 1$. Let the waiting time of the arrival be w', and the waiting time of the previous joiner be w. Then $w' = \max(0, w+v-u)$, where u and v are respectively the interarrival and service times of persons who previously joined. We can write

$$H_1(x) = \Pr(w+v-u \leqslant x)$$

whether or not $x = 0$; therefore

$$H_1(x) = \Pr(w \leqslant x+y)$$

where $y = u-v$, and for fixed y this is

$$H_1(x|y) = H_J(x+y) \tag{19}$$

since by hypothesis the distribution of w is $H_J(x)$. The distribution of y is easy to find, and turns out to be

$$f(y) = \begin{cases} \lambda\Delta\, e^{-\lambda y} & (y > 0), \\ \lambda\Delta\, e^{\mu y} & (y < 0), \end{cases}$$

† This formula can also be used to show that $H_J(\infty) = 1$.

where $\Delta = 1/(1+\rho)$. Applying this to (19),

$$H_1(x) = \int_{-x}^{0} \lambda\Delta\, e^{\mu y} H_J(x+y)\, dy + \int_0^\infty \lambda\Delta\, e^{-\lambda y} H_J(x+y)\, dy. \tag{20}$$

However

$$\int_0^\infty e^{-\lambda y} H_J(x+y)\, dy = \int_x^\infty e^{-\lambda(z-x)} H_J(z)\, dz = \int_0^\infty e^{-\lambda(z-x)} H_J(z)\, dz - \int_0^x e^{-\lambda(z-x)} H_J(z)\, dz$$

$$= e^{\lambda x} H_J^*(\lambda) - \int_0^x e^{\lambda(x-z)} H_J(z)\, dz.$$

Substituting this for the second integral in (20), taking the Laplace transform and using (5), we obtain

$$h_1^*(s) = \frac{\Delta}{1-\xi}[h^*(s)-p(0)] + \frac{s}{s-\lambda}\frac{\eta(\Delta)-p(0)}{\rho(1-\xi)} - \frac{\mu+s}{s-\lambda}\frac{\Delta}{1-\xi}[h^*(s)-p(0)] \quad (0 < s < \lambda). \tag{21}$$

It will be noticed that since $\Delta < 1$, $\eta(\Delta)$ is convergent. Replacing s by $\mu(1-s)/s$, we have

$$\eta_1(s) = \frac{\eta(s)-p(0)}{1-\xi}\frac{s\Delta}{s-\Delta} - \frac{\eta(\Delta)-p(0)}{1-\xi}\frac{\Delta(1-s)}{\rho(s-\Delta)} \quad (0 < \Delta < s < 1). \tag{22}$$

Expanding this expression gives

$$\left.\begin{aligned} p_1(0) &= \frac{\eta(\Delta)-p(0)}{\rho(1-\xi)}, \\ p_1(n) &= \frac{(1+\rho)^{n-1}}{1-\xi}\sum_{j=n}^{\infty}\Delta^j p(j) \quad (n > 0). \end{aligned}\right\} \tag{23}$$

Formulae (23) are consistent with the classical (no balking) case. For, if ξ be set equal to zero, then $p(n) = (1-\rho)\rho^n$, and $p_1(n)$ computed from (23) turns out also to have that value, as would be expected.

For values of $N > 1$, this method of waiting times must be replaced by another argument, based on the joint distribution of N and queue length, n. It appears that neither system will yield results in both the cases $N = 1$ and $N > 1$.

We denote the joint probability of a queue length of n and $N-1$ previous balkers at the time of a customer's arrival by $\pi(n, N)$, $n = 0, 1, \ldots$, $N = 1, 2, \ldots$, and the distribution of N alone by $\pi(N) = \sum_n \pi(n, N)$. Thus

$$\pi(n, N) = p_N(n)\,\pi(N). \tag{24}$$

Various relationships between the $\pi(n, N)$ can be found by setting up differential difference equations as in Feller (1950) or QWB, with the additional complication that the probabilities involve two variables. The initial values of n and N must be dealt with separately, and we have therefore four cases: Case I, $n = 0$, $N = 1$; Case II, $n = 0$, $N > 1$; Case III, $N = 1$, $n > 0$; Case IV, $n > 0$, $N > 1$. In each case there will be terms arising from the three possibilities: no transition, one arrival, one departure, if we omit to mention terms of $o(\delta t)$. The following discussion indicates briefly how each of the four equations is obtained; the time argument, which is almost immediately removed by assuming equilibrium, is momentarily indicated by a subscript.

Case I ($N = 1$, $n = 0$)

no transition	$(1-\lambda\delta t)\,\pi_t(0,1)$,
arrival	impossible,
departure	$\mu\delta t\,\pi_t(1,1)\,(1-\lambda\delta t)$

leading to the equation $\mu\pi(1,1)-\lambda\pi(0,1)=0.$ (25)

Case II ($n = 0$, $N > 1$)

no transition	$(1-\lambda\delta t)\,\pi_t(0,N)$,
arrival	impossible,
departure	$\mu\delta t(1-\lambda\delta t)\,\pi_t(1,N)$

leading to the equation $\mu\pi(1,N)-\lambda\pi(0,N)=0.$ (26)

Case III ($n > 0$, $N = 1$)

no transition	$(1-\lambda\delta t)\,(1-\mu\delta t)\,\pi_t(n,1)$,
arrival	$\lambda\delta t(1-\mu\delta t)\,G(n-2)\,p(n-1)$

since the arrival must join.

Departure	$\mu\delta t(1-\lambda\delta t)\,\pi_t(n+1,1)$

leading to the equation

$$\mu\pi(n+1,1)+\lambda G(n-2)\,p(n-1)-(\lambda+\mu)\,\pi(n,1)=0. \tag{27}$$

Case IV ($N > 1$, $n > 0$)

no transition	$(1-\lambda\delta t)\,(1-\mu\delta t)\,\pi_t(n,N)$,
arrival	$\lambda\delta t(1-\mu\delta t)\,F(n-1)\,\pi_t(n,N-1)$

since the arrival must balk.

Departure	$\mu\delta t(1-\lambda\delta t)\,\pi_t(n+1,N)$

leading to the equation

$$\mu\pi(n+1,N)+\lambda F(n-1)\,\pi(n,N-1)-(\lambda+\mu)\,\pi(n,N)=0. \tag{28}$$

The bivariate generating functions will be defined similarly to the univariate ones

$$\left.\begin{aligned}\eta(s,S)&=\Sigma\Sigma\pi(n,N)\,s^nS^N,\\ \xi(s,S)&=\Sigma\Sigma\pi(n+1,N)\,F(n)\,s^nS^N.\end{aligned}\right\} \tag{29}$$

Multiplying equations (25)–(28) by appropriate powers of s and S and summing gives the bivariate generating function equation

$$\mu S[\eta(s)-p(0)]+\lambda sS\xi(s,S)+\eta(s,S)\left[\frac{\mu}{s}-\lambda-\mu\right]+\left(\mu-\frac{\mu}{s}\right)\sum_{N=1}^{\infty}\pi(0,N)\,S^N=0. \tag{30}$$

If we set $s = 1$ or $S = 1$ in (30), we find relations equivalent to QWB (7); with both $s = S = 1$, we obtain $\xi(1,1) = \xi$.

A more useful (conditional) generating function equation can be obtained by using (24) in equations (26) and (28), for $N > 1$. Multiplying by s only, and summing on n only, gives

$$\lambda s^2\pi(N-1)\,\xi_{N-1}(s)-[(\lambda+\mu)\,s-\mu]\,\pi(N)\,\eta_N(s)-\mu(1-s)\,\pi(0,N)=0. \tag{31}$$

The expressions $\xi_N(1)$ are by no means equal to ξ; they will be abbreviated ξ_N. When we set $s = 1$ in (31), we find that the relationship between the probabilities $\pi(N)$ and the ξ_N is exactly the same as that found in QWB to exist between the $p(n)$ and the $\rho G(n-1)$, making allowance for the fact that the domain of N begins with unity rather than zero.

Namely, $$\pi(N+1) = \xi_N \pi(N) \quad (N > 1). \tag{32}$$

For $N = 1$, the equation which completes the set (31) is

$$(\Delta - s)\pi(1) + (1-s)\pi(0,1) + \Delta[\eta(s) - p(0)] = 0 \tag{33}$$

and with $s = 1$, this yields $$\pi(1) = 1 - \xi \tag{34}$$

which provides an initial value for the set (32). In fact, defining

$$C_1 = 1, \quad C_N = \prod_{i=1}^{N-1} \xi_i \quad (N > 1), \tag{35}$$

we can write $$\pi(N) = (1-\xi) C_N, \tag{36}$$

a result perfectly analogous to QWB (3). Thus the bivariate probabilities $\pi(n, N)$ have essentially the same type of marginal total distributions.

The $\pi(N)$ are known once the ξ_N are known. To discover these latter probabilities, we first set $s = \Delta$ in equation (31)

$$\Delta\pi(N-1)\xi_{N-1}(\Delta) = \pi(0,N) = p_N(0)\pi(N) = \xi_{N-1}\pi(N-1)p_N(0), \tag{37}$$

and therefore $$p_N(0) = \frac{\Delta\xi_{N-1}(\Delta)}{\xi_{N-1}} \quad (N > 1). \tag{38}$$

Putting the value for $\pi(0, N)$ given by (37) back into equation (31), we are able to express $\eta_N(s)$ in terms of various values of $\xi_{N-1}(s)$, namely

$$\eta_N(s) = \frac{(1-s)\Delta^2\xi_{N-1}(\Delta) - \Delta s^2 \xi_{N-1}(s)}{(\Delta - s)\xi_{N-1}}. \tag{39}$$

When $s < \Delta$, this can be expanded in positive powers of s, and the values of $p_N(n)$ abstracted from the series. In this way the value given in equation (38) is continued as follows

$$p_N(n) = \frac{\rho(1+\rho)^{n-1}}{\xi_{N-1}} \left\{\Delta\xi_{N-1}(\Delta) - \sum_{i=1}^{n-1} F(i-1)\, p_{N-1}(i)\, \Delta^i\right\} \quad (n > 1,\ N > 1). \tag{40}$$

Furthermore, since $$\xi_N = \sum_1 p_N(n) F(n-1)$$

and $$(1/s)[\eta_N(s) - p_N(0)] = \sum_1 p_N(n) s^{n-1},$$

values of ξ_N can be obtained by replacing s^n by $F(n)$ in an expansion of $(1/s)[\eta_N(s) - p_N(0)]$ in positive powers of s. When these are carried out, the following recursive formula is obtained

$$\xi_N = (\rho/\xi_{N-1})\left\{\Delta\xi_{N-1}(\Delta)\left[1 + \sum_{i=1}^{\infty} F(i)(1+\rho)^i\right] - \sum_{i=1}^{\infty}\sum_{j=1}^{i} F(i)\, p_{N-1}(j)\, F(j-1)(1+\rho)^{i-j}\right\}. \tag{41}$$

3. Interjoining gap distribution

In the preceding section we have found a method for calculation of $\pi(N)$, the probability that the time gap between arrival (whether joining or not) and preceding joiner will be

$$\frac{\lambda^N x^{N-1} e^{-\lambda x}}{\Gamma(N)}. \tag{42}$$

Therefore the density function of times between arrivals and preceding joiners is

$$a(x) = \sum_{N=1}^{\infty} \pi(N) \frac{\lambda^N x^{N-1} e^{-\lambda x}}{\Gamma(N)}. \tag{43}$$

In this section we find corresponding quantities for gaps between a *joiner* and the preceding joiner.

The reason for studying interjoining gap distributions can be simply stated. In actual observation of a queue with balking, it could happen that balkers would not be observed. For example, if the queue occurred at a heavily congested road intersection, the drivers who choose to 'go around the block' rather than wait would very likely adopt this plan at some distance from the intersection, and so be indistinguishable from ordinary traffic stream attrition. In such a case the problem of estimating the magnitude or quality of balking (which might be called seeking hidden demand for service) would have to be solved by reference to joiners only. Specifically, the observed interjoining gap distribution would be compared with the negative exponential on one hand, and with the distribution which we obtain in this section on the other.

Let the functions $\phi(s)$ and $\psi(s)$ be defined by

$$\phi(s) = \pi(1) + s\pi(2) + s^2\pi(3) + \ldots, \tag{44}$$

$$\psi(s) = \pi(1) + s\pi(2) + \frac{s^2}{2!}\pi(3) + \ldots. \tag{45}$$

Then there is a simple relationship

$$a(x) = \lambda e^{-\lambda x} \psi(\lambda x) \tag{46}$$

between the ψ function and the gap density. Also, it is easy to express ψ in terms of ϕ. For, with a gamma-type integral substituted for the factorials in (44), we have

$$\phi(s) = \int_0^{\infty} e^{-t} \psi(st)\, dt, \tag{47}$$

a formula mentioned by Riordan (1958, p. 29). Replacing s by $1/s$ and t by st gives

$$(1/s)\,\phi(1/s) = \int_0^{\infty} e^{-st} \psi(t)\, dt.$$

Therefore ψ is the inverse Laplace transform of $(1/s)\,\phi(1/s)$.

Now we consider the interjoining gap distribution, and define similar quantities. Let the probability that an interjoining gap consists of N negative exponential phases be $\gamma(N)$, and let $b(x)$ be the interjoining gap density function which we seek. Then, defining

$$\Phi(s) = \gamma(1) + s\gamma(2) + s^2\gamma(3) + \ldots, \tag{48}$$

$$\Psi(s) = \gamma(1) + s\gamma(2) + \frac{s^2}{2!}\gamma(3) + \ldots \tag{49}$$

we know already that $$b(x) = \lambda e^{-\lambda x}\Psi(\lambda x) \tag{50}$$

and that $\Psi(s)$ is the inverse Laplace transform of

$$(1/s)\,\Phi(1/s) = \Sigma\gamma(N)\,s^{-N-1}. \tag{51}$$

Since both ϕ and ψ are known in principle from the preceding section $b(x)$ will be known once $\Phi(s)$ is expressed in terms of $\phi(s)$. The required formula is

$$\Phi(s) = \frac{(s-1)\,\phi(s)}{s(1-\xi)} + \frac{1}{s} \tag{52}$$

which we shall now prove.

Consider a long sequence of mixed J and B symbols, and let

J_N = total number of J runs of length N,

B_N = total number of B runs of length N,

Π_N = total number of times a symbol is separated from the next preceding J by exactly $N-1$ symbols,

Γ_N = total number of times a J is separated from the next preceding J by exactly $N-1$ symbols.

Then† $$\pi(N) = \frac{\Pi_N}{\Sigma\Pi_N}, \quad \gamma(N) = \frac{\Gamma_N}{\Sigma\Gamma_N}, \quad j_N = \frac{J_N}{\Sigma J_N}, \quad b_N = \frac{B_N}{\Sigma B_N}$$

represent the relative frequencies. Note that $\Sigma J_N = \Sigma B_N$. We shall use the unorthodox notation $\bar{\pi}$, $\bar{\gamma}$, $\bar{\jmath}$, $\bar{b}$ for the means of these distributions to avoid the introduction of four additional letters. It is easy to see that

$$\bar{\gamma} = \frac{1}{1-\xi}. \tag{53}$$

Each run of N J's will correspond to N single π gaps and $N-1$ single γ gaps. Therefore

$$\Pi_1 = \Sigma N J_N \quad \text{and} \quad \Gamma_1 = \sum_2 (N-1)\,J_N. \tag{54}$$

On the other hand, a run of N B's produces one double π gap, one triple π gap, ..., one $(N+1)$-fold π gap; but only one $(N+1)$-fold γ gap. Hence

$$\Pi_N = \sum_{N+1} B_N, \quad \Gamma_N = B_{N-1} \quad (N = 2, 3, \ldots). \tag{55}$$

Summing the Π_N, we have

$$\begin{aligned}\Sigma\Pi_N &= \Sigma N J_N + \Sigma N B_N = \bar{\jmath}\Sigma J_N + \bar{b}\Sigma B_N \\ &= \pi(1)\Sigma\Pi_N + \bar{b}\pi(2)\Sigma\Pi_N,\end{aligned}$$

or $$\bar{b} = \frac{1-\pi(1)}{\pi(2)}. \tag{56}$$

Also, since $$\Sigma\Pi_N = (\bar{\jmath}+\bar{b})\Sigma B_N = (\bar{\jmath}+\bar{b})\pi(2)\Sigma\Pi_N,$$

we have $$\bar{\jmath} = \pi(1)/\pi(2). \tag{57}$$

Summing the Γ_N, $$\Sigma\Gamma_N = \sum_2 N J_N - \sum_2 J_N + \Sigma B_N = \pi(1)\Sigma\Pi_N; \tag{58}$$

and similarly $$\Gamma_N = B_{N-1} = \Pi_N - \Pi_{N+1} = [\pi(N) - \pi(N+1)]\Sigma\Pi_N. \tag{59}$$

† All summations in this section go to infinity; those without lower index begin at one.

Therefore
$$\gamma(N) = \frac{\pi(N) - \pi(N+1)}{\pi(1)}, \tag{60}$$

and since
$$\Sigma N\Gamma_N = \Sigma\Pi_N = \overline{\gamma}\Sigma\Gamma_N + \pi(1)\,\overline{\gamma}\Sigma\Pi_N$$

we have also
$$\pi(1) = 1/\overline{\gamma} = 1 - \xi. \tag{61}$$

Forming generating functions from (60) and using (61) with (53), it is now easy to deduce (52). The distribution of lengths of runs of balks is very similar to (60)

$$b_N = \frac{\gamma(N+1)}{1-\gamma(1)} = \bar{\jmath}\gamma(N+1). \tag{62}$$

Hence
$$\bar{b} = \bar{\jmath}\frac{\xi}{1-\xi}. \tag{63}$$

The mean number of phases in a π gap is connected in a very simple way with the variance of the number of phases in a γ gap, as may be seen if (60) is multiplied by N^2 and summed. The result is

$$(1-\xi)\,\mathrm{var}\,(\gamma) + \frac{1}{1-\xi} = 2\overline{\pi} - 1. \tag{64}$$

The preceding results are of course valid for any sequence of symbols of two sorts, whether arising from a queue or not, provided the proper construction is placed on the quantities involved. An example based on the following sequence may be useful:

BJJJBBBJBJBJBBJJJJBJJJJBJJJJJJJJBJJJJJBBJJBJBBJJJJ

N	1	2	3	4	5	6	7	8
J	4	1	1	3	1	0	0	1
j	4/11	1/11	1/11	3/11	1/11	0	0	1/11
B	7	3	1	—	—	—	—	—
b	7/11	3/11	1/11	—	—	—	—	—
Π	34	11	4	1	—	—	—	—
π	0·68	0·22	0·08	0·02	—	—	—	—
Γ	23	7	3	1	—	—	—	—
γ	0·6764	0·2059	0·0882	0·0294	—	—	—	—

$\bar{\jmath} = 3{\cdot}0909$, $\bar{b} = 1{\cdot}4545$, $\overline{\pi} = 1{\cdot}44$, $\overline{\gamma} = 1{\cdot}4704$, $\xi = 0{\cdot}32$, $1/(1-\xi) = 1{\cdot}4704$.

4. POISSON BALKING

In this section we present some numerical results based on a simulated queue with balking, in which 500 arrivals with $\lambda = 1, \mu = 2$, use a Poisson balking rule with mean value 10. That is

$$G(n) = \frac{\gamma(n+1, 10)}{\Gamma(n+1)}. \tag{65}$$

Table 1 shows the observed values for arrivals and Table 2 for joiners, together with fitted distributions (i) for geometric probabilities

$$(1-Q)\,Q^N \quad (N = 0, 1, \ldots), \tag{66}$$

(ii) for negative binomial probabilities (cf. Kendall, 1943, p. 125)

$$\left[\frac{C}{C+1}\right]^P \binom{P+N-1}{N} (C+1)^{-N} \quad (N = 0, 1, \ldots) \tag{67}$$

and (iii) for Borel probabilities (cf. Borel, 1942, Haight & Breuer, 1960)

$$\frac{N^{N-2}}{(N-1)!}e^{-\alpha N}\alpha^{N-1} \quad (N = 1, 2, \ldots). \tag{68}$$

The first two of these must refer to the number of previous balkers, while (68) refers directly to the number of exponential phases. We shall use a dash to indicate that the index is thus retarded one unit: $\pi'(N) = \pi(N+1)$, $\gamma'(N) = \gamma(N+1)$.

Table 1. *Observed and theoretical distributions of number of negative exponential phases in an interarrival gap*

N	$\pi(N)$ observed	$\pi'(N)$ geometric	$\pi'(N)$ negative binomial	$\pi(N)$ Borel
1	0·50928	0·45118	0·50016	0·57695
2	·21649	·24762	·21570	·18308
3	·10515	·13590	·11613	·08714
4	·05979	·07458	·06667	·04916
5	·03918	·04093	·03946	·03047
6	0·02474	0·02246	0·02378	0·02005
7	·01856	·01233	·01450	·01375
8	·01031	·00677	·00892	·00972
9	·00825	·00371	·00552	·00703
10	·00412	·00204	·00343	·00519
>10	·00412	0·0248	·00573	·01746
χ^2	—	17·20	1·92	16·01

Table 2. *Observed and theoretical distributions of number of negative exponential phases in an interjoining gap*

N	$\pi(N)$ observed	$\pi'(N)$ geometric	$\pi'(N)$ negative binomial	$\pi(N)$ Borel
1	0·57490	0·50928	0·58869	0·61263
2	·21862	·24991	·18943	·18390
3	·08907	·12264	·09356	·08281
4	·04049	·06018	·05158	·04419
5	·02834	·02953	·03029	·02591
6	0·01215	0·01449	0·01786	0·01613
7	·01619	·00711	·01087	·01046
8	·00405	·00349	·00671	·00700
9	·00809	·00171	·00418	·00479
10	·00405	·00084	·00262	·00334
>10	·00405	·00082	·00421	·00884
χ^2	—	10·39	3·92	6·21

The geometric and Borel distributions have been fitted by equating means (which is equivalent to maximum likelihood in these cases) and the negative binomial according to the equations given by Kendall.

The geometric distribution gives a very bad fit, and this refutes any suggestion that balking occurs at random in this case (cf. Feller, 1950, p. 219). Furthermore, from equation (52) or (60), if one of these distributions were geometric, the other would have to be geometric also.

The negative binomial distribution might be justified in this problem as a 'contagious' distribution, but I have not been able to establish a connexion. One difficulty seems to be that the parameters P and C do not have any intuitively simple relationship with important queue-theoretic quantities such as ξ.

In the Borel case, this difficulty does not exist, for $\alpha = \xi$ exactly, and therefore estimation of the Borel parameter is equivalent to estimation of the probability of balking. Furthermore, the interpretation of the Borel distribution as the probability that N persons will pass through a queue before it first vanishes, beginning with one person (random arrival and regular service) seems to have some vague connexion with the problem at hand.

I wish to express my gratitude to the referee for his elaborate and valuable comments on earlier versions of this paper, which exposed a number of serious errors.

REFERENCES

BOREL, E. (1942). Sur l'emploi du théorème de Bernoulli pour faciliter le calcul d'un infinité de coefficients. Application au problème de l'attente à un guichet. *C.R. Acad. Sci., Paris*, **214**, 452–6.

FELLER, W. (1950). *An Introduction to Probability Theory and its Applications*, **1**. New York: John Wiley and Sons.

GAVER, D. P. (1954). The influence of servicing time in queueing processes. *J. Op. Res. Soc. Am.* **2**, 139.

HAIGHT, F. A. (1957). Queueing with balking. *Biometrika*, **44**, 360–9.

HAIGHT, F. A. & BREUER, M. A. (1960). The Borel–Tanner distribution. *Biometrika*, **47**, 143–50.

KENDALL, M. G. (1943). *The Advanced Theory of Statistics*, **1**. London: Charles Griffin and Company.

RIORDAN, J. (1958). *An Introduction to Combinatorial Analysis*. New York: John Wiley and Sons.

TRUXAL, J. G. (1955). *Automatic Feedback Control System Synthesis*. New York: McGraw-Hill.

Biometrika (1960), **47**, 3 *and* 4, *p.* 297
With 4 text-figures
Printed in Great Britain

On comparing different tests of the same hypothesis*

BY H. A. DAVID AND CARMEN A. PEREZ
Virginia Polytechnic Institute

1. INTRODUCTION

Suppose that the standard test of the null hypothesis $H_0: \theta = \theta_0$ is based on the statistic v. In certain circumstances we may prefer a test involving some other statistics v', be it because such a test is more easily and quickly carried out or because it is more robust (i.e. insensitive to assumptions not under test). It is customary to compare the long-run individual performances of the two tests by means of their power functions or related criteria. While this kind of comparison is invaluable the following question also appears to us of interest. To what extent, in the long run, does the quick test give the same verdict as the standard test when both are applied to the same sets of data? In special cases this question has been studied by empirical methods. Thus agreement between the standard deviation s and the range w in a series of small samples of real data has been examined by Pearson & Haines (1935) who plotted w against s in a control chart situation. With the help of the known standard deviation σ they were able to insert on their diagrams a grid of several upper- and lower-percentage points of both w and s so that the corresponding degrees of significance attained by any sample and the agreement in this respect could be seen on inspection. A similar approach, but based on experimental sampling from normal populations, was used by Eisenhart & Swed (1940) to compare two possible tests of the homogeneity of several estimates of variance. In the null case they found considerable disagreement at the 5 % level of significance.

More recently Cox (1956) has discussed some general aspects of this problem. We now restate and add to some of his arguments: (*a*) The case for a particular quick test is strengthened if it is known that it will nearly always lead to the same action as the standard test. (*b*) Quick tests are often used in preliminary or spot analyses. It seems highly desirable that the conclusions so reached should not be seriously upset if a fuller analysis is carried out. The quick test should inspire confidence in this respect even if it will not be followed up at all. (*c*) The method of data collection or experimentation may lend itself to a particularly advantageous use of quick tests; indeed, the experimental approach may be modified for this reason. To be specific let us consider a life test on items with a normal distribution, $N(\theta, \sigma^2)$, of lives. The test may be stopped as soon as just over half the items in it have failed. If σ^2 is assumed known a very simple test on θ may be based on the sample median. One could, of course, improve the *analysis* by using all the available observations and not only the median. More fundamentally, the *experimental method* hinges on the fact that it allows a quick decision to be reached. Confidence in the reliability of this approach is presumably based on the knowledge that the median leads to a reasonably powerful test, but also on the supposition that the conclusion reached with the censored test would generally be confirmed were the test in fact run to completion. A similar point could be made in comparing the paired t-test and the sign test: not only does the sign test rest on fewer assumptions, it also allows a simplification of experimental technique.

* Research supported, in part, by Office of Ordnance Research, U.S. Army.

So far we have confined ourselves to consideration of a single quick test. If a competing test statistic v'' is also available it would seem reasonable to prefer the test based on v' to that using v'' if the former agrees more closely with the standard test, other things (e.g. power) being roughly equal. When both null and alternative hypotheses are simple this situation is illustrated schematically in Fig. 1 where the rejection regions are the sides of the boundaries shown which do not contain the origin.

In his article Cox studied the regression of v on v' in a special case and could then gauge the approximate degree of significance of v given the level of significance of v'. To do this he had to make some assumption about the prior distribution of θ which he took to be uniform. It is the purpose of the present paper to propose two measures of the degree of agreement between standard and quick tests which under suitable conditions do not have to rely on the

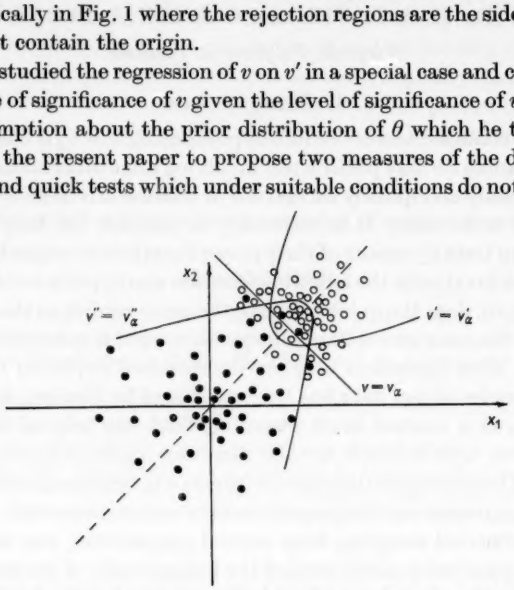


Fig. 1. Illustrating that of two equally powerful tests based on v', v'' the former provides better agreement with the best test based on v; ●, represents the probability density $f(x_1, x_2 | H_0)$, ○, represents the probability density $f(x_1, x_2 | H_1)$.

use of prior probabilities. These conditions hold in a number of important situations and require the existence of a complete sufficient statistic for θ if H_0 is simple and for θ as well as other unknown parameters if H_0 is composite (see e.g. Lehmann & Scheffé, 1950).

The two measures will be introduced by means of a simple example and then applied to give comparisons of:

(i) tests of location in a normal population with known variance;

(ii) the use of standard deviation, range and mean deviation in tests of dispersion for a normal parent;

(iii) the paired t-test and the sign test in the normal case.

It will also be shown that in large samples the two measures can be related to the asymptotic efficiencies (when these exist) of the statistics used in the quick tests or to the asymptotic relative efficiencies of these tests. A general statement of the two approaches introduced is given in §5.

2. Tests of location

We consider the following very simple situation: $x_1, x_2, \ldots, x_n$ is a random sample from an $N(\mu, 1)$ population. To test $H_0: \mu = 0$ against $H_1: \mu > 0$. Here the uniformly most powerful test is, of course, to reject H_0 when

$$u_1 = \bar{x}\sqrt{n} > u_\alpha,$$

where u_α is the upper α significance point of a unit normal deviate u. A possible quick test is to use only the first k ($< n$) observations in the sample and to reject H_0 if, with obvious notation,

$$u_2 = \bar{x}_k \sqrt{k} > u_\alpha.$$

One natural way of measuring the degree of agreement of the two tests is to obtain

$$\Pr(u_1 > u_\alpha \,|\, u_2 > u_\alpha),$$

the conditional probability of establishing significance at level α with u_1, given that the quick test u_2 has led to significance at that level. Now this probability is readily evaluated on H_0 or for any specified value of μ, but we would like to know its value for μ unknown. This can only be done if μ is assumed to have some prior distribution.

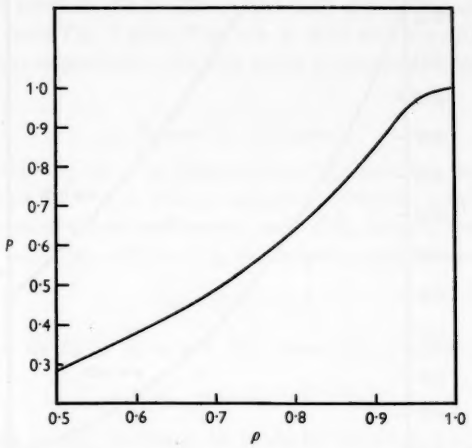


Fig. 2. $P = \Pr(u_2 > 1{\cdot}645 \,|\, u_1 = 2{\cdot}326)$ as a function of ρ, the correlation coefficient of the unit normal variates u_1, u_2.

The merits of u_1 and u_2 may, however, be compared differently. We propose two quite general approaches and illustrate them on the present example. Imagine a series of samples of size n each on the borderline of significance when u_1 is the test-statistic. The true value of μ is unknown and may be different from sample to sample. How would the quick test have worked on these samples?

(*a*) Consider the probability P of establishing significance with u_2 at level β, given that u_1 is *just* significant at level α, viz.

$$P = \Pr(u_2 > u_\beta \,|\, u_1 = u_\alpha).$$

Note that P does not depend on μ since u_1 is sufficient for μ. Since $P < \frac{1}{2}$ for $\alpha = \beta$ it is most meaningful to choose $\beta > \alpha$ (i.e. $u_\beta < u_\alpha$). We have

$$P = \Pr[u > (u_\beta - \rho u_\alpha)/\sqrt{(1-\rho^2)}], \tag{1}$$

where $\rho = \sqrt{(k/n)}$. A slightly different possibility is to find u_β (and hence β) such that P has a specified value.

Equation (1) will, of course, continue to hold when u_2 is any unit normal deviate having correlation coefficient ρ with u_1. In Fig. 2, P is plotted as a function of ρ for $\alpha = 0{\cdot}01$, $\beta = 0{\cdot}05$.

(*b*) An alternative approach arises from the question: Given u_1 is just significant at level α how significant is the *expected value* of u_2 calculated from the same sample? Since

$$\mathscr{E}(u_2 \,|\, u_1 = u_\alpha) = \rho u_\alpha,$$

the required level of significance is

$$\gamma = \Pr(u_2 > \rho u_\alpha). \tag{2}$$

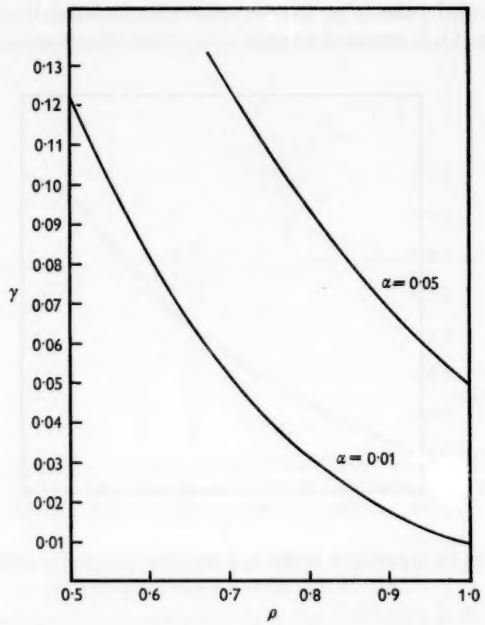


Fig. 3. Equivalent Type I error γ of quick test based on u_2 corresponding to significance level α of best test based on u_1: $\gamma = \Pr(u_2 > \rho u_\alpha)$

It is clear that $\gamma > \alpha$. We shall term γ the *equivalent Type I error* of the quick test based on u_2. Fig. 3 shows γ as a function of ρ for $\alpha = 0{\cdot}05$ and $0{\cdot}01$. The measure γ is further discussed in §5.

Generalizations

Consider the above situation but with $\bar{x}_k$ replaced by any unbiased estimator m of μ, such as the median, mid-point (mid-range), or some linear function of the order statistics. P may now be written as

$$P = \Pr(m > m_\beta \,|\, \bar{x} = \bar{x}_\alpha), \quad (\bar{x}_\alpha = u_\alpha/\sqrt{n}).$$

As is well known, $m - \bar{x}$ has a distribution free of μ and is statistically independent of $\bar{x}$, a result which follows most readily from the sufficiency and completeness of $\bar{x}$ for μ (Basu, 1955). Hence

$$\begin{aligned} P &= \Pr(m - \bar{x} > m_\beta - \bar{x} \,|\, \bar{x} = \bar{x}_\alpha) \\ &= \Pr(m - \bar{x} > m_\beta - \bar{x}_\alpha). \end{aligned}$$

This is liable to be rather awkward to evaluate and the second approach is somewhat simpler since

$$\mathscr{E}(m \mid \bar{x} = \bar{x}_\alpha) = \mathscr{E}(m - \bar{x} + \bar{x} \mid \bar{x} = \bar{x}_\alpha)$$
$$= \mathscr{E}(m - \bar{x}) + \bar{x}_\alpha = \bar{x}_\alpha.$$

Hence the 'equivalent Type I error' of m is

$$\gamma = \Pr(m > \bar{x}_\alpha).$$

It is of interest to examine the case where m is asymptotically normal with asymptotic efficiency E_m. In this situation Figs. 2 and 3 can be used at once to evaluate the large sample values of P and γ by using the relation $\rho = \sqrt{E_m}$ (of course, this relation is true in finite samples also, with E_m taken to be $1/(n \operatorname{var} m)$, but is not easily interpreted unless the joint distribution of $\bar{x}$ and m is bivariate normal). As an illustration take m to be the median. Then $E_m = 2/\pi$ and $\rho = 0{\cdot}798$ which from Fig. 2 gives P as low as 0·64 and from Fig. 3 yields $\gamma = 0{\cdot}094$, 0·031 for $\alpha = 0{\cdot}05$, 0·01, respectively. The mid-point is not asymptotically normal (Gumbel, 1958, p. 311).

3. Tests of dispersion

Given a random sample $x_1, x_2, \ldots, x_n$ from a $N(\mu, \sigma^2)$ population, suppose it is required to test the null hypothesis $H_0: \sigma = 1$, with μ unknown. We shall apply the two approaches introduced above to compare three well-known tests of H_0, namely, tests based on the sample standard deviation $s = \sqrt{[\Sigma(x_i - \bar{x})^2/(n-1)]}$, the range w, and the mean deviation

$$d = \Sigma |x_i - \bar{x}|/n.$$

When the alternative to H_0 is $H_1: \sigma > 1$ the three tests consist in rejecting H_0 when, respectively,

$$s > s_\alpha, \quad w > w_\alpha, \quad d > d_\alpha.$$

The first of these tests would, of course, be made by referring $\Sigma(x_i - \bar{x})^2$ to tables of the percentage points of χ^2. Percentage points of w and d are conveniently given in *Biometrika Tables for Statisticians*, **1** (Tables 22, 21). It may be mentioned that in this case also $\rho(s, w) = \sqrt{E_w}$ and $\rho(s, d) = \sqrt{E_d}$. These correlation coefficients have been tabulated by Hartley (1955) for $n = 2(1)20$. We have

$$P_w \equiv \Pr(w > w_\beta \mid s = s_\alpha) = \Pr\left(\frac{w}{s} > \frac{w_\beta}{s} \mid s = s_\alpha\right)$$
$$= \Pr\left(\frac{w}{s} > \frac{w_\beta}{s_\alpha}\right),$$

since w/s is independent of s, a result used by David *et al.* (1954) to find percentage points of this ratio of the range to the standard deviation calculated from the same sample.* For $n \geqslant 10$, P_w can be evaluated by a Pearson Type approximation to w/s. For $n = 10$, $\alpha = 0{\cdot}01$, $\beta = 0{\cdot}05$ we find $P_w = 0{\cdot}809$, and by exactly the same approach $P_d = 0{\cdot}905$. The corresponding correlation coefficients are 0·954 and 0·922. It is interesting to note that for $n = 3$ (and the same α, β) both P_w and P_d are unity. This follows from the fact that both w/s

* The result can be proved by elementary methods but follows most readily from Basu's (1955) work referred to in §2.

and d/s are bounded, with lower bounds respectively greater than w_β/s_α and d_β/s_α. (The lower bounds are $\sqrt{3}$ and $\frac{2}{3}$ while $w_\beta/s_\alpha = 1{\cdot}542$ and $d_\beta/s_\alpha = 0{\cdot}595$.)

The equivalent Type I errors γ_w, γ_d are readily obtained; thus

$$\begin{aligned}\mathscr{E}(w \mid s = s_\alpha) &= \mathscr{E}\left(\frac{w}{s} s \mid s = s_\alpha\right) \\ &= s_\alpha \mathscr{E}\left(\frac{w}{s} \mid s = s_\alpha\right) \\ &= s_\alpha \mathscr{E}(w/s) \\ &= s_\alpha \mathscr{E}(w)/\mathscr{E}(s).*\end{aligned}$$

The second step is justified by the properties of conditional expectations (see e.g. Kolmogorov, 1950, p. 56), and the last step is a consequence of the independence of w/s and s. We have, therefore,

$$\gamma_w = \Pr[w > s_\alpha \mathscr{E}(w)/\mathscr{E}(s)]$$

which may be evaluated from tables of the probability integral of w (Table 23 of *Biometrika Tables for Statisticians*, **1**). Likewise

$$\gamma_d = \Pr[d > s_\alpha \mathscr{E}(d)/\mathscr{E}(s)]$$

can be obtained from tables of the probability integral of d (Godwin & Hartley, 1945). In Table 1 we give, for $n \leqslant 10$ and $\alpha = 0{\cdot}05$, $0{\cdot}01$ the values of γ_w and γ_d together with the corresponding correlation coefficients taken from Hartley's (1955) paper. It will be seen that on all counts w is superior to d for $n = 4$, 5 but inferior for $n > 6$. The correlation coefficient agrees remarkably well with γ as an indicator of the relative merits of w and d. The results for $n = \infty$ follow from the asymptotic normality of d which has asymptotic efficiency $1/(\pi - 2)$.

Table 1. *Values of the 'equivalent Type I errors' γ_w, γ_d, of tests based on the range w and the mean deviation d corresponding to Type I error α in standard (χ^2) test; together with the correlation coefficients $\rho(s, w)$, $\rho(s, d)$*

n \ α	0·05 γ_w	0·01 γ_w	$\rho(w, s)$	0·05 γ_d	0·01 γ_d	$\rho(d, s)$
3	0·0508	0·0105	0·9959	0·0508	0·0105	0·9959
4	·0526	·0114	·9875	·0536	·0122	·9818
5	·0548	·0125	·9771	·0557	·0130	·9726
6	·0571	·0136	·9660	·0569	·0136	·9664
7	·0594	·0149	·9546	·0578	·0139	·9620
8	·0617	·0161	·9434	·0585	·0141	·9586
9	·0641	·0174	·9324	·0589	·0143	·9561
10	·0664	·0186	·9219	·0593	·0145	·9540
∞	—	—	—	·0618	·0147	·9359

4. THE SIGN TEST AND THE t-TEST

When a random sample $x_1, x_2, \ldots, x_n$ from an $N(\mu, \sigma^2)$ population is available and σ^2 is unknown, the uniformly most powerful test of $H_0: \mu = 0$ against $H_1: \mu > 0$ is to reject H_0 if

$$t = \bar{x}\sqrt{n}/s > t_\alpha,$$

* This result proves incidentally that the regression of w on s is linear. A similar argument shows that the variance of w given s is proportional to s^2.

where t_α is the upper α significance point of a t-variate with $n-1$ degrees of freedom. To apply the sign test to this situation the number r of positive x's is first determined. On H_0 (and in fact under a much more general null hypothesis) r is a binomial variate with parameters $(n, \frac{1}{2})$ so that a test of significance is easily made.

We begin by obtaining the conditional probability π_c that $x_1 > 0$, given that the t-test is just significant at level α.

$$\begin{aligned}\pi_c &= \Pr(x_1 > 0 \,|\, t = t_\alpha) \\ &= \Pr(x_1/s > 0 \,|\, t = t_\alpha) \\ &= \Pr\left(\frac{x_1-\bar{x}}{s} + \frac{\bar{x}}{s} > 0 \,\middle|\, t = t_\alpha\right).\end{aligned}$$

The distribution of $(x_1-\bar{x})/s$ is clearly free of μ and σ. Therefore $(x_1-\bar{x})/s$ is independent of any function of the complete sufficient statistic $(\bar{x}, s)$, and in particular of t. It follows that

$$\pi_c = \Pr\left(\frac{x_1-\bar{x}}{s} > -\frac{t_\alpha}{\sqrt{n}}\right).$$

This is then the conditional probability of a 'success' on the first trial and does not depend on μ. The n trials are evidently not independent so that the probability of r successes out of n is generally difficult to evaluate. However, the conditional probability of n straight successes can be expressed as

$$P' \equiv \Pr\left(\frac{x_{\min.}-\bar{x}}{s} > \frac{-t_\alpha}{\sqrt{n}}\right) = \Pr\left(\frac{x_{\max.}-\bar{x}}{s} < \frac{t_\alpha}{\sqrt{n}}\right).$$

The ratio $(x_{\max.}-\bar{x})/s$ was first studied by Pearson & Chandra Sekar (1936). With the help of further recent work by Borenius (1958) we obtain the following results:

n	Significance level of sign test	P' $\alpha = 0{\cdot}05$	P' $\alpha = 0{\cdot}01$
4	1/16	0·359	1
5	1/32	0·074	0·809
6	1/64	0·055	0·417

Our second approach is again easier to apply. After some manipulation we have

$$\left.\begin{aligned}\pi_c &= \Pr\left[t < t_\alpha \sqrt{\frac{n-2}{(n-1)^2 - t_\alpha^2}}\right] && (n-1)^2 - t_\alpha^2 > 0, \\ &= 1 && (n-1)^2 - t_\alpha^2 \leqslant 0,\end{aligned}\right\} \tag{3}$$

where t (but not t_α) is based on $n-2$ degrees of freedom. Hence the equivalent Type I error of the sign test is

$$\gamma = \Pr(r > n\pi_c), \tag{4}$$

where r is binomial $(n, \frac{1}{2})$; or with a normal approximation to r

$$\gamma \doteqdot \Pr[u > \sqrt{n}(2\pi_c - 1)]. \tag{5}$$

Since, in general, $n\pi_c$ is not integral we have thought it most appropriate to evaluate γ in (4) from the histogram representation of the binomial distribution; i.e. denoting the integral part of $n\pi_c + \frac{1}{2}$ by I, we take

$$\gamma = \Pr(r \geqslant I+1) + (n\pi_c + \tfrac{1}{2} - I)\Pr(r = I).$$

Table 2 gives γ for various values of α and $\nu = n-2$. Due to the discreteness of the binomial distribution some irregularities in the variation of γ with ν are inevitable.

Table 2. *Equivalent Type I error γ of the sign test corresponding to the paired t-test in samples of n with Type I error α*

$\nu = n-2$ \ α	0·05	0·025	0·01	0·005
2	0·0582	0·0312	0·0312	0·0312
3	·0918	·0308	·0171	·0156
4	·0892	·0522	·0145	·0103
5	·0912	·0501	·0240	·0077
6	·0971	·0518	·0245	·0142
7	·0872	·0570	·0211	·0138
8	·0972	·0511	·0278	·0119
9	·0945	·0574	·0262	·0161
10	·0939	·0568	·0264	·0152
12	·0896	·0587	·0263	·0171
15	·0956	·0577	·0271	·0174
20	·0964	·0576	·0281	·0182
30	·0950	·0565	·0301	·0187
60	·0950	·0588	·0308	·0193
∞	·0947	·0589	·0317	·0199

Relation to asymptotic relative efficiency (A.R.E.)

Stuart (1954) has pointed out that the asymptotic relative efficiency (in the sense of Pitman) of two consistent *tests* V, V' of the same null hypothesis concerning the parameter θ is equal to the ratio of the asymptotic variances of two consistent *estimators* v, v' of θ on which these tests can be based. Suppose now that v is sufficient for θ. Then it will be asymptotically efficient (provided an asymptotically efficient estimator of θ exists). In this case A.R.E. of V' relative to V will be equal to the asymptotic efficiency, say E', of v'. Asymptotically, if v and v' are jointly normal they will have correlation coefficient $\rho = \sqrt{E'}$. Thus we can apply the results and graphs of §2 to convert a known A.R.E. into values of P or γ. Conversely, knowing P or γ for two asymptotically normal statistics, one of which is sufficient, we can deduce their correlation coefficient ρ and hence the A.R.E. of the inefficient test. For example, it can be shown that as $n \to \infty$, equation (4) becomes

$$\gamma = \Pr[u > u_\alpha \sqrt{(2/\pi)}],$$

so that $\rho = \sqrt{(2/\pi)}$ on comparison with equation (2). This in turn yields the known A.R.E. of the sign test for a normal parent, viz. $2/\pi$.

5. Summary and discussion

We now restate the problem under consideration in its general form: Given two or more tests V, V', V'', ... of the same hypothesis $H_0: \theta = \theta_0$ against the alternative $H_1: \theta > \theta_0$, to find measures of the degree to which the tests lead to the same conclusion. Suppose that V, V', V'', ... are based respectively on the statistics v, v', v'', ... which have critical upper α significance points v_α, v'_α, v''_α, ... One measure of the degree of agreement of v and v' is

$$P = \Pr(v' > v'_\beta \mid v = v_\alpha) \quad (\beta > \alpha). \tag{6}$$

We regard P as useful if a statistic v can be found such that P does not depend on θ, which will be so if the conditional distribution of v' given v is free of θ. In this situation a second measure is the 'equivalent Type I error' γ of V', relative to V, defined by

$$\gamma = \Pr[v' > \mathscr{E}(v' \mid v = v_\alpha)]. \tag{7}$$

The following way of looking at γ may be of interest. The conditional distribution $f(v' \mid v = v_\alpha)$ has, for a reasonable test-statistic v', considerably smaller variance than the unconditional null distribution $f(v')$. Fig. 4 shows the two distributions in a typical case (with v' roughly the range in normal samples of 10 and v the corresponding standard deviation). For samples with $v = v_\alpha$ we may expect large values of v'. Suppose that for one of these samples $v' = v'_1$. This corresponds to a level of significance on the quick test of $\Pr(v' > v'_1)$. As an average measure of this level it is convenient to take γ as in (7). (Other averages might be used but, in view of the relatively small variance of $f(v' \mid v = v_\alpha)$, would not lead to very different results.)

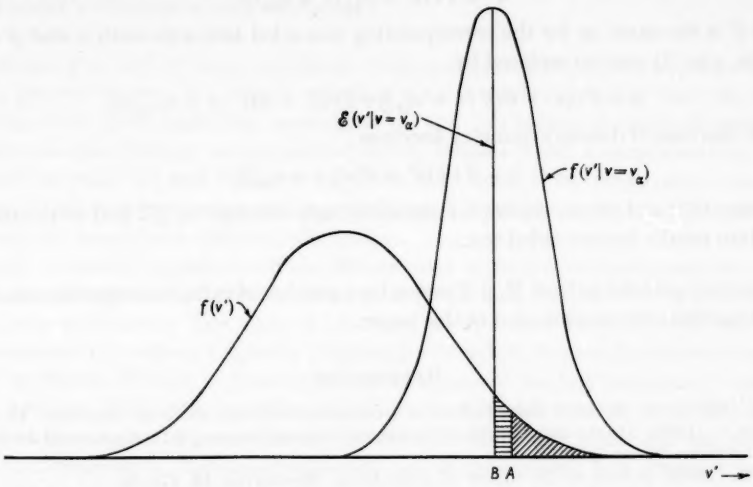


Fig. 4. Shaded area to right of A represents critical region for test based on v'; shaded area to right of B represents equivalent Type I error γ of the test.

It will be seen that the statistic v plays a special role in these measures. For H_0 a simple hypothesis, P and γ are free of θ if v is a sufficient statistic for θ. This case is illustrated in § 2. When H_0 is composite, suppose that the unspecified parameter (possibly vector-valued) is ϕ. If a statistic (v, y) exists which is sufficient for (θ, ϕ) then

$$\Pr(v' > v'_\beta \mid v, y)$$

is independent of θ. If, in addition, y is statistically independent of v' then P and γ will not depend on θ. The tests in § 3 fall into this category, with $v = s$ and $y = \bar{x}$. It is easy to see that tests V'', V''', ... may be compared with V in the same way. Also indirect comparisons of V' and V'' may be made with the help of V by evaluating P and γ for both V' and V''. In § 4, H_0 is composite and it is possible to find a satisfactory statistic v (viz. t) without the existence of y.

It seems to us that P and γ are intuitively reasonable measures of agreement but an alternative interpretation is possible. $\Pr(v' > v'_\beta \mid v)$ is, in all but artificial cases, a monotonic function of v, which we may without loss of generality take to be non-decreasing. Thus P may be interpreted as a lower bound of

$$\Pr(v' > v'_\beta \mid v \geqslant v_\alpha),$$

although this probability depends on θ. Likewise γ is a lower bound of

$$\Pr[v' > \mathscr{E}(v' \mid v \geqslant v_\alpha)].$$

For the sake of simplicity we have restricted our treatment to the case of one-sided tests, but this is not a necessary limitation. When the alternative hypothesis is two-sided (6) can be replaced by $P = \frac{1}{2}[\Pr(v' > v'_{\frac{1}{2}\beta} \mid v = v_{\frac{1}{2}\alpha}) + \Pr(v' < v'_{1-\frac{1}{2}\beta} \mid v = v_{1-\frac{1}{2}\alpha})]$.

If, in particular, both v and v' have symmetrical distributions (the case of 'double symmetry') this reduces to

$$P = \Pr(v' > v'_{\frac{1}{2}\beta} \mid v = v_{\frac{1}{2}\alpha}), \tag{8}$$

so that P is the same as for the corresponding one-sided test with both α and β halved. Likewise, γ in (7) may be replaced by

$$\gamma = \Pr[v' > \mathscr{E}(v' \mid v = v_{\frac{1}{2}\alpha})] + \Pr[v' < \mathscr{E}(v' \mid v = v_{1-\frac{1}{2}\alpha})],$$

which in the case of double symmetry becomes

$$\gamma = 2\Pr[v' > \mathscr{E}(v' \mid v = v_{\frac{1}{2}\alpha})]. \tag{9}$$

By means of (8) and (9) the results for one-sided tests obtained in §§ 2 and 4 are easily converted into results for two-sided tests.

We are very grateful to Prof. E. S. Pearson for a number of valuable suggestions, including Fig. 1, regarding the presentation of this paper.

REFERENCES

BASU, D. (1955). On statistics independent of a complete sufficient statistic. *Sankhyā*, **15**, 377–80.

BORENIUS, G. (1958). On the distribution of the extreme values in a sample from a normal distribution. *Skand. Aktuar. Tidskr.* 131–66.

COX, D. R. (1956). A note on the theory of quick tests. *Biometrika*, **43**, 478–80.

DAVID, H. A., HARTLEY, H. O. & PEARSON, E. S. (1954). The distribution of the ratio, in a single normal sample, of range to standard deviation. *Biometrika*, **41**, 482–93.

EISENHART, C. & SWED, FRIEDA S. (1940). On certain criteria for testing the homogeneity of k estimates of variance (Abstract). *Ann. Math. Statist.* **11**, 111–12.

GODWIN, H. J. & HARTLEY, H. O. (1945). Probability integral and percentage points of the mean deviation in normal samples. *Biometrika*, **33**, 254–65.

GUMBEL, E. J. (1958). *Statistics of Extremes*. New York: Columbia University Press.

HARTLEY, H. O. (1955). Some recent developments in analysis of variance. *Comm. Pure Appl. Math.* **8**, 47–72.

KOLMOGOROV, A. N. (1950). *Foundations of the Theory of Probability*. New York: Chelsea Publishing Company.

LEHMANN, E. L. & SCHEFFÉ, H. (1950). Completeness, similar regions and unbiased estimation. Part I. *Sankhyā*, **10**, 305–40.

PEARSON, E. S. & HAINES, JOAN (1935). The use of range in place of standard deviation in small samples. *J.R. Statist. Soc.* (Suppl.) **2**, 83–98.

PEARSON, E. S. & CHANDRA SEKAR, C. (1936). The efficiency of statistical tools and a criterion for the rejection of outlying observations. *Biometrika*, **28**, 308–20.

STUART, A. (1954). Asymptotic relative efficiencies of distribution-free tests of randomness against normal alternatives. *J. Amer. Statist. Ass.* **49**, 147–57.

Biometrika (1960), **47**, 3 *and* 4, *p.* 307
Printed in Great Britain

The performance of some correlation coefficients for a general bivariate distribution

By D. J. G. FARLIE

Research Laboratories of The General Electric Company Limited, Wembley, England

1. Introduction

Association between two variables has in the past been measured or tested by several coefficients, among them

(i) Product moment correlation coefficient.
(ii) Spearman's rank correlation coefficient.
(iii) Kendall's alternative rank coefficient τ.
(iv) Probability of concordance.

The efficiency of each of these coefficients in detecting association between two variables is not in general known, but it is known that none is uniformly better than the others. Conjectures have been made that coefficients (ii) and (iii) behave in much the same way for reasonable distributions, but as pointed out by Daniels (1950) a sample ranking can be considered for which (ii) and (iii) have very different values. If the population is such that rankings of this type have appreciable probability, it would be expected that tests based on (ii) and (iii) would have different efficiencies.

It would obviously be desirable if the performance of these coefficients could be studied for a fairly general class of distributions, in order to gain insight into the relationship between the coefficients. The class of bivariate distributions obtained from monotonic transformations of variates originally possessing a bivariate Normal distribution has been studied by Fieller, Hartley & Pearson (1957) by means of random samples of correlated Normal deviates. They concluded that for sample sizes 30 and 50 with correlation greater than 0·6 in the original bivariate normal distribution Kendall's coefficient was probably better than Spearman's. Correspondence with one of the authors, E. S. Pearson, indicates that the Fisher-Yates statistic mentioned in their paper is better than either rank correlation coefficient within this same field. Another such general class of distributions has been considered by Konijn (1958), namely, the class derived from linear combinations of two independent variables.

It is the purpose of this paper first to propose a general class of bivariate distribution functions; secondly, to study how the various coefficients of association compare in efficiency with each other and with a maximum likelihood estimator of the parameter of association; and finally to use the general results to determine which are the best coefficients to use in a number of special cases. The efficiency measure used in this paper will be the asymptotic relative efficiency, in the neighbourhood of independence, as defined by Pitman (1948).

The form of the bivariate distribution function proposed in this paper is an extension of an idea of Morgenstern; see, for example, Gumbel (1958). Morgenstern (1956) proposed $F(x)\,G(y)\,[1+\alpha\{1-F(x)\}\{1-G(y)\}]$ as a bivariate distribution function having $F(x)$, $G(y)$ as marginal distribution functions, and Gumbel noted that this differed from the bivariate normal distribution if $F(x) \equiv \Phi(x)$ and $G(y) \equiv \Phi(y)$ where Φ is the normal error function.

2. PROPOSED CLASS OF ALTERNATIVE HYPOTHESES

Let $F(x)$ be the marginal continuous distribution function of x,

$G(y)$ be the marginal continuous distribution function of y,

$A\{F(x)\}$ be a function of x that tends to zero as $F(x) \to 1$,

$B\{G(y)\}$ be a function of y that tends to zero as $G(y) \to 1$.

Then $H(x,y) = F(x)\,G(y)\,[1+\alpha A(F)\,B(G)]$ is a suitably general class of bivariate distribution functions for which the marginal distribution functions are $F(x)$ and $G(y)$.

The functions, $A(F)$ and $B(G)$, are not completely arbitrary but, as will be shown later, it will suffice if we choose for $A(F)$ and $B(G)$, functions that are bounded and have bounded first differential coefficients with respect to their arguments. Since α is a parameter measuring association, we can arrange without loss of generality that the least upper bounds on $|A(F)|$ and on $|B(G)|$ are both 1. The expression for $H(x,y)$ is then unique.

Consider a general family of two-dimensional distribution functions, $L(x,y;\alpha)$, depending on a parameter α, which tend smoothly to $F(x)\,G(y)$, a distribution function for independent variables, as $\alpha \to 0$. Then for sufficiently small α

$$\begin{aligned} L(x,y;\alpha) &\equiv F(x)\,G(y) + \alpha \frac{\partial L}{\partial \alpha}(x,y;\alpha)\Big|_{\alpha=0} + o(\alpha) \\ &= F(x)\,G(y) + \alpha K(x,y)\,F(x)\{1-F(x)\}\,G(y)\{1-G(y)\} + o(\alpha), \quad \text{say.} \end{aligned}$$

Now $K(x,y)$ may be expanded as a series of terms of the type $a_i x^{(i)} y^{(i)}$ where $x^{(i)}$ and $y^{(i)}$ are sets of orthogonal functions defined on $F(x)$, $G(y)$ respectively so that

$$\int x^{(i)} x^{(j)}\,dF(x) = \int y^{(i)} y^{(j)}\,dG(y) = \delta_{ij}.$$

There are many such expansions: for example, it is possible to expand $K(x,y)$ in a double sine series with $\frac{1}{2}\{\pi F(x)\}$ and $\frac{1}{2}\{\pi G(y)\}$ as the arguments; expansion in polynomials, such as Legendre polynomials, is also possible. Whatever method we use will provide an expansion of the type $H(x,y) = F(x)\,G(y)\{1+\alpha\Sigma a_i A_i(F)\,B_i(G)\}$, where $A_i(F)$, $B_i(G)$ tend to zero as $F(x)$ and $G(y)$ tend to 1 by virtue of the factors $(1-F)$, $(1-G)$, and so are of the type considered in this paper.

We shall in this paper work with the simple form $H(x,y) = F(x)\,G(y)\{1+\alpha A(F)\,B(G)\}$ described above, but in fact the main results extend immediately to the more general form and hence can be regarded as applying to any family of bivariate distribution functions that tend smoothly to independence as $\alpha \to 0$.

3. CONDITIONS FOR $H(x,y)$ TO BE A CONTINUOUS BIVARIATE DISTRIBUTION FUNCTION

For $H(x,y)$ to be a continuous bivariate distribution function the following must hold:

(1) $H(x,y)$ is a differentiable monotonic increasing function of both x and y.

(2) $0 \leqslant H(x,y) \leqslant 1$ for all x and y.

(3) $\lim\limits_{x,\,y\to\infty} H(x,y) = 1$, $\lim\limits_{x,\,y\to-\infty} H(x,y) = 0$.

For $H(x,y)$ to have marginal distribution functions $F(x)$ and $G(y)$ it must further satisfy

$$\lim_{x\to\infty} H(x,y) = G(y) \quad \text{and} \quad \lim_{y\to\infty} H(x,y) = F(x).$$

Consider the proposed form of $H(x,y)$, namely $H(x,y) = F(x)\,G(y)\{1+\alpha A(F)\,B(G)\}$. Now.

$$\lim_{x,y\to\infty} H(x,y) = F(\infty)\,G(\infty)+\alpha F(\infty)\,G(\infty)\,A(1)\,B(1)$$
$$= 1+\alpha A(1)\,B(1) = 1.$$

Since $A(1)$ and $B(1)$ are both assumed to be zero. Further,

$$\lim_{x,y\to-\infty} H(x,y) = 0$$

because A and B are both finite. Now we require that $H(x,y)$ should be a monotonic increasing function of x and y; i.e. that

$$\frac{\partial^2 H}{\partial x\,\partial y} = F'(x)\,G'(y)\left\{1+\alpha\frac{d(FA)}{dF}\frac{d(GB)}{dG}\right\} \geqslant 0.$$

Since FA is a continuous function of F that is zero at both ends of the interval $(0,1)$ and we may suppose that $FA \not\equiv 0$, the derivative $\{d(FA)\}/dF$ must attain both negative and positive values, and similarly for $\{d(GB)\}/dG$: the product of the two can also, therefore, have both positive and negative values. We have required the derivatives to be bounded so that $-M_1 \leqslant \{d(FA)\}/dF \leqslant M_2$ and $-M_3 \leqslant \{d(GB)\}/dG \leqslant M_4$, where M_1, M_2, M_3, M_4 are chosen to be the least values for which these inequalities hold. The product

$$\frac{d(FA)}{dF}\frac{d(GB)}{dG}$$

may vary as follows:

$$-\max(M_2M_3, M_1M_4) \leqslant \frac{d(FA)}{dF}\frac{d(GB)}{dG} \leqslant \max(M_1M_3, M_2M_4);$$

call these two limits $-M'$ and $+M''$. Then

$$1+\alpha\frac{d(FA)}{dF}\frac{d(GB)}{dG} \geqslant 0$$

whenever

$$\alpha\frac{d(FA)}{dF}\frac{d(GB)}{dG} \geqslant -1;$$

that is, when $\alpha \geqslant -1/M''$ and $\alpha \leqslant +1/M'$. When this condition holds, we know that $H(x,y)$ is monotonic, increasing from 0 to 1, and therefore must satisfy condition (1). The marginal distribution functions are to be $F(x)$ and $G(y)$, now

$$\lim_{x\to\infty} H(x,y) = F(\infty)\,G(y)\{1+\alpha A(1)\,B(y)\}$$
$$= G(y), \quad \text{since} \quad A(1) = 0.$$

Similarly

$$\lim_{y\to\infty} H(x,y) = F(x)$$

because $B(1) = 0$.

4. Regression of Y on X

When two variables, X and Y, are not independent, the mean value taken by one of them, say X, for a fixed value of the other, say $Y = y$, is not usually a constant independent of y. The relationship between this conditional expectation and the fixed value y, the regression equation, is then of special interest. Considering our general bivariate function

$$H(x,y) = F(x)\,G(y)\{1+\alpha A(x)\,B(y)\}$$

we have,
$$E(X|Y=y)=\int_{-\infty}^{+\infty} x\frac{\partial^2 H}{\partial x\,\partial y}\left(\frac{dG}{dy}\right)^{-1}dx$$
$$=\int_{-\infty}^{+\infty} x\,dF(x)+\alpha\frac{d(GB)}{dG}\int_{-\infty}^{+\infty} x\frac{d(FA)}{dx}dx$$

Now $\int_{-\infty}^{+\infty} x\frac{d(FA)}{dx}dx$ is a constant, k say, so that the regression equation is

$$E(X|Y=y)=\mu_x+\alpha k\frac{d(GB)}{dG},$$

where μ_x is the unconditional mean of x. The only restrictions on $\{d(GB)\}/dG$ are

(i) $GB=0$ for $G=0$ and $G=1$, (ii) $|GB|$ is bounded, (iii) $\left|\frac{d(GB)}{dG}\right|$ is bounded.

Over a suitable range of α, therefore, we may make the regression of x on y any shape we please and examples are quoted later of linear, quadratic, triangular and discontinuous regression lines.

4·1. *Examples*

Linear regression equations

(*a*) Let y be rectangularly distributed between 0 and 1, so that

$$G(y)=y \quad (0\leqslant y\leqslant 1)$$

and put $B=1-G$. Then

$$E(X|Y=y)=\mu_x+(1-2y)\int_{-\infty}^{+\infty} x\frac{d(FA)}{dx}dx,$$

so that the regression of x on y is linear.

(*b*) In general, let Y have the distribution function G with mean μ_y and variance σ_y^2, and put $\{d(GB)\}/dG=(Y-\mu_y)$. Then $GB=\int_0^G (Y-\mu_y)\,dG$ satisfies $GB\to 0$ as $G\to 0$ and also $GB\to 0$ as $G\to 1$. If Y is of finite range then $\{d(GB)\}/dG$ and GB will both be finite and there will be a range of α for which the bivariate function is a true distribution function. If Y is not of finite range then there will be no range of α for which the bivariate function is a true distribution function. However, we may arrange that our general bivariate function represents a truncated form of the distribution. Choose the truncation points, X_1, X_2, Y_1, Y_2, such that

$$X_1=\mu_x-\sqrt{(M/\epsilon)},\quad Y_1=\mu_y-\sqrt{(M/\epsilon)},$$
$$X_2=\mu_x+\sqrt{(M/\epsilon)},\quad Y_2=\mu_y+\sqrt{(M/\epsilon)};$$

then our bivariate function is a true distribution function for values of α between $\pm\epsilon/M$. Asymptotic relative efficiency is measured against alternatives that differ from the null hypothesis by amounts that become vanishingly small in the limit and we can choose these alternatives to be of the form $\alpha=an^{-\frac{1}{2}}$, where a is a constant. This value of α will be inside the range $\pm\epsilon/M$ if we make ϵ a function of n such that $an^{-\frac{1}{2}}<\epsilon(n)/M<bn^{-\frac{1}{2}}$ say. Then in our limiting process not only does the parameter of association tend to zero but the truncation introduced cuts off a proportion of the original distribution that tends to zero at least as quickly as $n^{-\frac{1}{2}}$ by virtue of Tchebychev inequality, $P_r\{(X-\mu)>\sqrt{(M/\epsilon)}\}\leqslant\sigma^2\epsilon/M$. Continuity considerations lead us to assign the same A.R.E. to both the true alternatives and the truncated alternatives in the neighbourhood of $\alpha=0$.

With this modification to the original bivariate function permitted, it is possible to represent the bivariate normal distribution function in the form

$$H(x,y) = F(x)\,G(y)\{1+\alpha A(x)\,B(y)\}.$$

Consider, $$H(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} \frac{1}{2\pi(1-\rho^2)^{\frac{1}{2}}} \exp -\tfrac{1}{2}(x^2-2\rho xy+y^2)\,dx\,dy$$

for small values of ρ and for x and y not too large: we can expand the integrand in powers of ρ and obtain

$$H(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} \frac{1}{2\pi}(\exp -\tfrac{1}{2}x^2)\,(\exp -\tfrac{1}{2}y^2)\,(1+\rho xy+o(\rho))\,dx\,dy$$

by ignoring the small probability that x and y take large values. This can be written as $H(x,y) = F(x)\,G(y)\{1+\alpha A(x)\,B(y)\}$ if we put

$$F(x) = \frac{1}{\sqrt{(2\pi)}}\int_{-\infty}^{x} \exp -\tfrac{1}{2}t^2\,dt, \quad G(y) = \frac{1}{\sqrt{(2\pi)}}\int_{-\infty}^{y} \exp -\tfrac{1}{2}t^2\,dt,$$

$$\frac{d(FA)}{dF} = x \quad \text{and} \quad \frac{d(GB)}{dG} = y.$$

Quadratic regression equations

(*a*) Let y be rectangularly distributed on the interval $(0, 1)$. Then,

$$G(y) = y \quad (0 \leqslant y \leqslant 1),$$

put $B(y) = 1-y^2$; then the regression equation becomes

$$E(X|\,Y=y) = \mu_x + (1-3y^2)\int_{-\infty}^{+\infty} x\frac{d(FA)}{dx}\,dx.$$

This is a quadratic regression but it corresponds to a monotonic regression line; it is possible, however, to make the expectation of x reach a maximum or minimum value in the range of y.

(*b*) Let y again be rectangularly distributed on the interval $(0, 1)$ and put $B = 1-3y+2y^2$,

$$\frac{d(GB)}{dG} = 1-6G+6G^2 = 1-6y+6y^2.$$

This is a quadratic expression that reaches its minimum value where $\{d^2(GB)\}/dG^2 = 0$; that is at $G = \frac{1}{2}$. Thus the regression equation

$$E(X|\,Y=y) = \mu_x + (1-6y+6y^2)\int_{-\infty}^{+\infty} x\frac{d(FA)}{dx}\,dx$$

has a stationary value at $y = \frac{1}{2}$.

Triangular regression equations

(*a*) Let y be rectangularly distributed on $(0, 1)$ and put

$$B = 1-2y \qquad (y \leqslant \tfrac{1}{2}),$$
$$B = 2y-3+(1/y) \quad (G \geqslant \tfrac{1}{2}),$$
$$B = 0 \qquad (G > 1,\ G < 0).$$

The regression equation then reduces to

$$E(X|\,Y=y) = \mu_x + (1-4y)\int_{-\infty}^{+\infty} x\frac{d(FA)}{dx}\,dx \quad (y \leqslant \tfrac{1}{2})$$

$$= \mu_x + (4y-3)\int_{-\infty}^{+\infty} x\frac{d(FA)}{dx}\,dx \quad (y \geqslant \tfrac{1}{2}).$$

Discontinuous regression equations

It is not, however, necessary that $\{d(GB)\}/dG$ should be continuous but only that GB should be continuous; as an example consider y to be rectangularly distributed on $(0, 1)$. Put

$$B = 1-2y \quad (G \leqslant \tfrac{1}{2}), \qquad B = 0 \quad (G > \tfrac{1}{2})$$

and

$$\frac{d(GB)}{dG} = 1-4y \quad (G \leqslant \tfrac{1}{2}), \qquad \frac{d(GB)}{dG} = 0 \quad (G > \tfrac{1}{2}).$$

The regression equation then reduces to the form

$$\begin{aligned} E(X|Y=y) &= \mu_x + \int_{-\infty}^{+\infty} x \frac{d(FA)}{dx}\, dx(1-4y) \quad (y \leqslant \tfrac{1}{2}) \\ &= \mu_x \qquad\qquad\qquad\qquad\qquad\quad (y > \tfrac{1}{2}) \end{aligned}$$

and is composed of two line segments that are disconnected.

When we compare the efficiency of two rank correlation coefficients with one another, the answer is clearly invariant under monotonic functional transformations of x or y. Therefore we may consider a canonical form of the distribution obtained after probability integral transformations have converted the marginal distributions into rectangular distributions over $(0, 1)$. Then the distribution function is

$$xy\{1+\alpha A(x)\, B(y)\}$$

and the regression equation of x on y is

$$\begin{aligned} E(X|Y=y) &= \mu_x + \alpha k \frac{d}{dy}\{yB(y)\} \\ &= \mu_x + \alpha k\{B(y) + yB'(y)\}. \end{aligned}$$

Thus the functions $A(x)$, $B(y)$ are very simply related to the slope of the regression curves in the canonical form. Subject to the restrictions on A, B considered above, regression equations of any slope can be produced by choice of A and B, and conversely the form of the regression curve determines A, B. If we wish to produce two distributions with the same regression lines but, say, different scedastic curves we must pass to the more general form described in § 2.

5. The evaluation of the mean value of Spearman's rank correlation coefficient

Consider the Spearman coefficient of rank correlation

$$r_s = 1 - \frac{6\sum_1^n (d_k^2)}{n(n^2-1)},$$

where d_k is the difference between the ranks of the x and y variates for the kth member of the sample.

To determine the asymptotic relative efficiency of the Spearman's rank correlation coefficient when compared with the maximum likelihood estimator of α it is necessary to know:

(*a*) the variance when $\alpha = 0$;

(*b*) the mean value when α is small.

The first of these is $1/(n-1)$ (Kendall, 1943). The mean value under the hypothesis of dependence will be calculated in the following discussion. First, it is necessary to define some symbols. The subject of study will be samples of n pairs of values, one x value and one y value in each pair (the two values might for instance be the hardness and the tensile strength of some test piece) and we can consider the pair as a single entry to form one element of the sample. We can name these n elements $1, 2, \ldots, n$ without considering the variate values composing the sample element, in such a way that each element is equally likely to receive any number from 1 to n. After this allocation of names we can refer to the x and y values associated with the kth sample element and we write

x_k to represent the x value associated with the kth element,

y_k to represent the y value associated with the kth element,

and $i(x_k)$ to be the number of x values in the sample less than or equal to x_k. Similarly $j(y_k)$ is the number of y values in the sample less than or equal to y_k. The functions $i(x_k)$ and $j(y_k)$ serve to rank the x and y variate values.

Let $\mathbf{i}$ be the column vector $\{i(x_1), i(x_2), \ldots, i(x_n)\}'$, and let $\mathbf{j}$ be the column vector $\{j(y_1), j(y_2), \ldots, j(y_n)\}'$.

The probability that the first element will have values (x_1, y_1), the second (x_2, y_2), etc., is

$$dP = \prod_{\kappa} \left\{1 + \alpha \frac{d(FA)}{dF(x_k)} \frac{d(GB)}{dG(y_k)}\right\} dF(x_k)\, dG(y_k).$$

For small values of α we may expand this as far as terms linear in α to give

$$dP = \left\{1 + \alpha \sum_{\kappa} \frac{d(FA)}{dF(x_k)} \frac{d(GB)}{dG(y_k)}\right\} \prod_{\kappa} dF(x_k)\, dG(y_k).$$

The probability that the two column vectors of ranks are $\mathbf{i}$ and $\mathbf{j}$ will then be given by integrating dP over the region R for which the ranks $i(x_k), j(y_k)$; $k = 1, 2, \ldots, n$ are the same as those observed.

The ordering of the names of the sample elements is immaterial to the determination of r_s; each of the orders is equally likely and there are $n!$ such orderings. The probability of any arrangement of x and y ranks will therefore be $n!$ times the probability just calculated for the occurrence of the two ranking vectors $\mathbf{i}$ and $\mathbf{j}$. The expected value of r_s is given by

$$E(r_s) = n! \, \Sigma r_s(\mathbf{i}, \mathbf{j}) \int_R \ldots \int \left\{1 + \alpha \sum_{\kappa} \frac{d(FA)}{dF(x_k)} \frac{d(GB)}{dG(y_k)}\right\} \prod_{\kappa} dF(x_k)\, dG(y_k),$$

where Σ is taken as the summation over all distinct arrangements of $\mathbf{i}, \mathbf{j}$. Two arrangements will be indistinguishable if they can be reached from one another by identical permutations of the rows of the $\mathbf{i}$ and $\mathbf{j}$ vectors.

We may ignore the unity term since this corresponds to $E(r_s)$ when $\alpha = 0$ and is therefore 0. Also

$$\Sigma n! \int_R \ldots \int \left\{\Sigma \frac{d(FA)}{dF(x_k)} \frac{d(GB)}{dG(y_k)}\right\} \prod_{\kappa} dF(x_k)\, dG(y_k) = 0,$$

since this represents the coefficient of α in the probability that some arrangement of x and y values occurs and this has a constant value of one.

Now by virtue of the equivalent form of Spearman's rank correlation coefficient as the product moment correlation coefficient of the ranks considered as variate values, we can write

$$r_s = \frac{12}{n(n^2-1)} \left[\mathbf{i}'\mathbf{j} - \frac{n(n+1)^2}{4}\right].$$

Therefore
$$E(r_s) = \frac{12}{n(n^2-1)}\left[E(\mathbf{i'j}) - \frac{n(n+1)^2}{4}\right]$$
$$= \frac{12}{n(n^2-1)}\, n!\,\alpha\, \Sigma\, \mathbf{i'j} \int_R \cdots \int \Sigma \frac{d(FA)}{dF(x_k)} \frac{d(GB)}{dG(y_k)}\, \mathbf{dF\, dG}.$$

Consider now one item in the inner sum; say, that associated with the kth sample element. Call this I_k. Then
$$I_k = \int_R \cdots \int \frac{d(FA)}{dF(x_k)} \prod_l dF(x_l) \frac{d(GB)}{dG(y_k)} \prod_i dG(y_i).$$

This is separable into two parts, one in x and one in y. Thus
$$I_k = \int_{R_1} \cdots \int \frac{d(FA)}{dF(x_k)} \prod_i dF(x_i) \int_{R_2} \cdots \int \frac{d(GB)}{dG(y_k)} \prod_i dG(y_i),$$

where R_1 is the region where the x's are ordered in the desired way and R_2 is the region where the y's are ordered in the desired way. Change the variables in the problem to F_i $(i = 1, \ldots, n)$ and G_j $(j = 1, \ldots, n)$ so that $F\{(x_k)\}$ is written as $F_i\{(x_k)\}$ and $G\{(y_k)\}$ is written as $G_j\{(y_k)\}$. The suffices i and j correspond to the value of the function $i(x_k)$ and $j(y_k)$; that is, the ranks of the x and y values associated with the kth sample element.

$$I_k = \int_{F_{n-1}}^1 dF_n \int_{F_{n-2}}^1 dF_{n-1} \cdots \int_{F_i}^1 dF_{i+1} \int_0^1 \frac{dF_i A(F_i)}{dF_i} \int_0^{F_i} dF_{i-1} \cdots \int_0^{F_2} dF_1$$
$$\times \int_{G_{n-1}}^1 dG_n \cdots \int_0^1 \frac{dG_j B(G_j)}{dG_j} \int_0^{G_j} dG_{j-1} \cdots \int_0^{G_2} dG_1$$
$$= \int_0^1 \frac{(1-F)^{n-i}}{(n-i)!} \frac{dFA(F)}{dF} \frac{F^{i-1}}{(i-1)!} dF \int_0^1 \frac{(1-G)^{n-j}}{(n-j)!} \frac{dGB(G)}{dG} \frac{G^{j-1}}{(j-1)!} dG$$
$$= \theta_k(i)\, \Phi_k(j).$$

The total inner sum is therefore obtained by summing, over k, these contributions $\theta_k(i)\, \Phi_k(j)$ and if we define
$$\boldsymbol{\theta} = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_n \end{bmatrix} \quad \text{and} \quad \boldsymbol{\Phi} = \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_n \end{bmatrix},$$

this sums $\sum_\kappa \theta_k(i)\, \Phi_k(j) = \boldsymbol{\theta'\Phi}$. Now
$$E(r_s) = \frac{12n!\,\alpha}{n(n^2-1)}\, \Sigma(\mathbf{i'j})\, (\boldsymbol{\theta'\Phi}),$$

of which the more interesting part is the summation. Call this sum S; then
$$S = \Sigma(\mathbf{i'j})\, (\boldsymbol{\theta'\Phi})$$

contains $n^2 n!$ terms that can be grouped into $n^2(n-1)^2\, n-2!$ terms of the type $kl\theta(i)\, \Phi(j)$, $i \neq k$, $j \neq l$; $n^2(n-1)\, n-2!$ terms of the type $ij\theta(i)\, \Phi(j)$, suffices to θ and Φ can now be abandoned.

Summation over the $k \neq i$ and the $l \neq j$ gives

$$S = \sum_i \sum_j \theta(i)\,\Phi(j)\left[\left(\frac{n(n+1)}{2} - i\right)\left(\frac{n(n+1)}{2} - j\right) + (n-1)\,ij\right](n-2)!$$

$$= \sum_i \sum_j \theta(i)\,\Phi(j)\left[\left(\frac{n(n+1)^2}{2}\right) - \frac{n(n+1)\,(i+j)}{2} + nij\right](n-2)!$$

$$= (n-2)!\left\{\sum_i \theta(i)\frac{n(n+1)}{2}\sum_j \Phi(j)\frac{n(n+1)}{2}\right.$$

$$\left. - \frac{n(n+1)}{2}\left[\sum_i i\theta(i)\sum_j \Phi(j) + \sum_j j\Phi(j)\sum_i \theta(i)\right] + n\sum_i i\theta(i)\sum_j j\Phi(j)\right\}.$$

But
$$\Sigma\theta(i) = \sum_{i=1}^{n}\int_0^1 \frac{(1-F)^{n-i}}{n-i!}\frac{F^{i-1}}{i-1!}\frac{dFA(F)}{dF}dF = \frac{1}{(n-1)!}\int_0^1 \frac{dFA(F)}{dF}dF$$

$$= \frac{1}{(n-1)!}[FA(F)]_0^1 = 0.$$

Similarly $\Sigma\Phi(j) = 0$.

$$\Sigma i\theta(i) = \Sigma(i-1+1)\,\theta(i) = \Sigma(i-1)\,\theta(i) \quad \text{by the preceding result}$$

$$= \Sigma\int_0^1 \frac{(1-F)^{n-i}}{n-i}\frac{F^{i-2}}{i-2!}\frac{Fd\,FA(F)}{dF}dF$$

$$= \frac{1}{n-2!}\int_0^1 F\frac{dFA(F)}{dF}dF.$$

Therefore,
$$S = \frac{(n-2)!\,n}{n-2!\,n-2!}\int_0^1 F\frac{dFA(F)}{dF}dF\int_0^1 G\frac{dGB(G)}{dG}dG,$$

so that
$$E(r_s) = \frac{12n!\,\alpha n}{n(n^2-1)\,(n-2)!}\int_0^1 F\frac{dFA(F)}{dF}dF\int_0^1 G\frac{dGB(G)}{dG}dG$$

$$= \frac{12\alpha n}{n+1}\left\{[F^2A]_0^1 - \int_0^1 FA\,dF\right\}\left\{[G^2B]_0^1 - \int_0^1 GB\,dG\right\}$$

$$= \frac{12\alpha n}{n+1}\left\{\int_0^1 FA\,dF\right\}\left\{\int_0^1 GB\,dG\right\} \to 12\alpha\left\{\int_0^1 FA\,dF\right\}\left\{\int_0^1 GB\,dG\right\}$$

as $n \to \infty$.

The estimate of α by the Spearman coefficient is therefore,

$$\alpha = \frac{r_s}{12\int_0^1 FA\,dF\int_0^1 GB\,dG}$$

for large values of n, and in the null case this estimate has variance

$$\frac{1}{n}\frac{1}{144\left\{\int_0^1 FA\,dF\right\}^2\left\{\int_0^1 GB\,dG\right\}^2}$$

for large values of n.

6. THE EVALUATION OF THE MEAN VALUE OF KENDALL'S ALTERNATIVE COEFFICIENT OF CORRELATION, t

6·1. *General*

Consider now the coefficient of correlation defined by

$$t = \frac{2S}{n(n-1)},$$

where S is the score obtained from a sample of n bivariate observations (x, y) by arranging either x or y, say x, in the natural order and considering the disorder produced in the y values. The magnitude of the score is found by taking the y values in pairs from this ordered sequence of observations and scoring $+1$ if the ordering is the same as in the natural ordering of the y values and scoring -1 if the ordering is the reverse of the natural ordering.

Hoeffding (1947) obtained the mean and variance of the distribution of this alternative coefficient by considering the probability of concordance. The probability of concordance, π, is defined as the probability that, if two elements of the bivariate population are chosen, say $(x_1, y_1)(x_2, y_2)$, then the product $(x_1 - x_2)(y_1 - y_2)$ is positive. With this definition, an estimate, p, of π obtained from the sample is obviously the number of concordant pairs out of the $\frac{1}{2}\{n(n-1)\}$ possible pairs divided by $\frac{1}{2}\{n(n-1)\}$ but the number of concordant pairs is just the number of $+1$'s in the score S. Therefore

$$p\frac{(n-1)n}{2} - (1-p)\frac{n(n-1)}{2} = S,$$

$$2S = 2pn(n-1) - (n-1)$$

and

$$t = 2p - 1;$$

also from Hoeffding, when $\alpha = 0$, the variance of $t = 4/(9n)$.

6·2. *Particular results*

The probability of concordance, π, is defined as the probability that, if two elements of the population are chosen, say $(x_1, y_1)(x_2, y_2)$, then $(x_1 - x_2)(y_1 - y_2) > 0$.

Now the probability of $(x_1, y_1)(x_2, y_2)$ is $h(x_1, y_1)\,h(x_2, y_2)\,dx_1 dx_2 dy_1 dy_2$ and the probability conditional upon $x_1 < x_2$ is $2h(x_1, y_1)\,h(x_2, y_2)\,dx_1 dx_2 dy_1 dy_2$.

The probability required is the integral of this probability element over the region $y_1 < y_2$; i.e.

$$\pi = 2\int_{-\infty}^{+\infty}\int_{-\infty}^{x_2}\int_{-\infty}^{+\infty}\int_{-\infty}^{y_2} h(x_1, y_1)\,h(x_2, y_2)\,dy_1 dy_2 dx_1 dx_2;$$

when α is small we may neglect terms in α^2 and higher powers.

$$\pi = 2\int_{-\infty}^{+\infty}\int_{-\infty}^{x_2}\int_{-\infty}^{+\infty}\int_{-\infty}^{y_2}\left\{1 + \alpha\left[\frac{dF_1 A_1}{dF_1}\frac{dG_1 B_1}{dG_1} + \frac{dF_2 A_2}{dF_2}\frac{dG_2 B_2}{dG_2}\right]\right\} dG_1 dG_2 dF_1 dF_2$$

changing the variables from x_1, x_2, y_1, y_2 to F_1, F_2, G_1, G_2,

$$\pi = 2\left\{(\tfrac{1}{2})^2 + \alpha\left[\int_0^1\int_0^{F_2}\frac{dF_1 A_1}{dF_1}dF_1 dF_2 \int_0^1\int_0^{G_2}\frac{dG_1 B_1}{dG_1}dG_1 dG_2 + \int_0^1\int_0^{F_2} dF_1 \frac{dF_2 A_2}{dF_2}dF_2 \int_0^1\int_0^{G_2} dG_1 \frac{dG_2 B_2}{dG_2}dG_2\right]\right\}$$

$$= \tfrac{1}{2} + 4\alpha\left\{\int_0^1 FA\,dF \int_0^1 GB\,dG\right\}.$$

Therefore
$$E(t) = 2E(p) - 1$$
$$= 2\pi - 1$$
$$= 8\alpha\left(\int_0^1 FA\,dF \int_0^1 GB\,dG\right).$$

The estimate of α by Kendall's coefficient, t, is

$$\alpha_2 = \frac{t}{8\int_0^1 FA\,dF \int_0^1 GB\,dG}$$

and in the null case this estimate has variance

$$\frac{4}{9n}\frac{1}{64\left\{\int_0^1 FA\,dF \int_0^1 GB\,dG\right\}^2},$$

or
$$\frac{1}{144n\left\{\int_0^1 FA\,dF \int_0^1 GB\,dG\right\}^2},$$

asymptotically the same as the estimate by Spearman's correlation coefficient. Therefore, these two rank correlation estimates have the same asymptotic relative efficiency.

This property holds for the more general bivariate function mentioned in § 2, namely

$$H(x, y) = F(x)\,G(y)\{1 + \alpha\Sigma a_i A_i(x)\,B_i(y)\}.$$

An important step in any of the arguments seeking the expected value of a statistic for small values of α is that which equates

$$\prod_l \left(1 + \alpha\frac{d(FA)}{dF_l}\frac{d(GB)}{dG_l}\right) dF_l dG_l$$

with
$$\left(1 + \alpha\sum_l \frac{d(FA)}{dF_l}\frac{d(GB)}{dG_l}\right)\prod_l dF_l dG_l.$$

This step for the more general bivariate function leads to

$$\left(1 + \alpha\sum_l\sum_i a_i \frac{d(FA)_i}{dF_l}\frac{d(GB)_i}{dG_l}\right)\prod_l dF_l dG_l$$

which can be written as

$$\left(1 + \alpha\sum_i a_i \sum_l \frac{d(FA)_i}{dF_l}\frac{d(GB)_i}{dG_l}\right)\prod_l dF_l dG_l.$$

It can be seen therefore that the expected value under this more general hypothesis is a weighted sum of the expected values under a series of hypotheses of the type already considered, the weight used being a property of the distribution function and not of the statistics being studied. We have shown that for all hypotheses of the simple type, $E(r_s)$ is $1{\cdot}5E(t)$ and this must also apply to similarly weighted sums of these expectations. From this it follows that the asymptotic relative efficiencies of the two rank correlation coefficients is the same for all distribution functions that allow an expansion of the more general kind in which each term separately could represent a true distribution function when taken with the term independent of α.

7. THE MEAN AND VARIANCE OF THE PRODUCT MOMENT CORRELATION COEFFICIENT

The variance of the product moment correlation coefficient

$$r = \frac{X_1}{\sqrt{(X_2 X_3)}},$$

where $X_1 = \Sigma XY - \frac{\Sigma X \Sigma Y}{n}$, $X_2 = \Sigma X^2 - \frac{(\Sigma X)^2}{n}$ and $X_3 = \Sigma Y^2 - \frac{(\Sigma Y)^2}{n}$

can be shown to be asymptotically $1/n$ (Kendall, 1943).

When n is large, X_1, X_2, X_3 will all be close to their respective mean values

$$\mu_1 = E(X-\mu_x)(Y-\mu_y), \quad \mu_2 = E(X-\mu_x)^2, \quad \mu_3 = E(Y-\mu_y)^2,$$

so that we may write

$$X_1 = \mu_1 + \epsilon_1, \quad X_2 = \mu_2(1+\epsilon_2), \quad X_3 = \mu_3(1+\epsilon_3),$$

where ϵ_1, ϵ_2 and ϵ_3 have zero mean and very small variances. Expanding in terms of powers of ϵ_1, ϵ_2, ϵ_3 gives

$$r \sim \frac{\mu_1}{\sqrt{(\mu_2\mu_3)}}(1 - \tfrac{1}{2}\epsilon_2 - \tfrac{1}{2}\epsilon_3) + \frac{\epsilon_1}{\sqrt{(\mu_2+\mu_3)}}.$$

Taking expectation of both sides gives

$$E(r) = E\frac{(X-\mu_x)(Y-\mu_y)}{\sigma_x\sigma_y} = \alpha \int_0^1 \left(\frac{X-\mu_x}{\sigma_x}\right)\frac{d(FA)}{dF}dF \int_0^1 \left(\frac{Y-\mu_y}{\sigma_y}\right)\frac{d(GB)}{dG}dG.$$

The estimator α_3, for α, obtained from the product moment correlation coefficient is, therefore,

$$\alpha_3 = \frac{r}{\int_0^1 \left(\frac{X-\mu_x}{\sigma_x}\right)\frac{d(FA)}{dF}dF \int_0^1 \left(\frac{Y-\mu_y}{\sigma_y}\right)\frac{d(GB)}{dG}dG}$$

and this estimator has in the null case a variance of

$$\frac{1}{n\left\{\int_0^1 \left(\frac{X-\mu_x}{\sigma_x}\right)\frac{d(FA)}{dF}dF \int_0^1 \left(\frac{Y-\mu_y}{\sigma_y}\right)\frac{d(GB)}{dG}dG\right\}^2}.$$

But

$$\int_0^1 \frac{\mu_x}{\sigma_x}\frac{d(FA)}{dF}dF = 0$$

so that this variance may be written as

$$\frac{\sigma_x^2\sigma_y^2}{n\left\{\int_0^1 X\frac{d(FA)}{dF}dF \int_0^1 Y\frac{d(GB)}{dG}dG\right\}^2}.$$

8. THE MAXIMUM LIKELIHOOD ESTIMATOR OF THE PARAMETER

The probability of obtaining $(x_1y_1)(x_2y_2)(x_3y_3)\ldots(x_ny_n)$ as the n sample elements is

$$\prod_i \left\{1 + \alpha\frac{d(FA)_i}{dF_i}\frac{d(GB)_i}{dG_i}\right\}dF_i dG_i.$$

The log-likelihood is, therefore,

$$L = \sum_i \log\left(1 + \alpha\frac{d(FA)_i}{dF_i}\frac{d(GB)_i}{dG_i}\right),$$

$$\frac{\partial L}{\partial \alpha} = \sum_i \frac{\frac{d(FA)_i}{dF_i}\frac{d(GB)_i}{dG_i}}{\left(1 + \alpha\frac{d(FA)_i}{dF_i}\frac{d(GB)_i}{dG_i}\right)}.$$

The maximum likelihood estimator is found by equating this to zero. When α is small and for large n, $\hat{\alpha}$ will also be small so that we may expand

$$\frac{1}{\left(1+\alpha\frac{d(FA)_i}{dF_i}\frac{d(GB)_i}{dG_i}\right)} \simeq 1-\alpha\frac{d(FA)_i}{dF_i}\frac{d(GB)_i}{dG_i}$$

so that the maximum likelihood equation becomes

$$\sum_i \frac{d(FA)_i}{dF_i}\frac{d(GB)_i}{dG_i} - \hat{\alpha}\sum_i\left(\frac{d(FA)_i}{dF_i}\frac{d(GB)_i}{dG_i}\right)^2 \simeq 0.$$

The variance of this estimator is obtainable from the form of the likelihood function and under suitable conditions the asymptotic variance of $\hat{\alpha}$ is given by

$$\operatorname{var}\hat{\alpha} = \left[\frac{-1}{nE(\partial^2 L/\partial\alpha^2)}\right]_{\alpha=0}.$$

Now
$$\left(\frac{\partial^2 L}{\partial\alpha^2}\right)_{\alpha=0} = -\left(\frac{d(FA)}{dF}\right)^2\left(\frac{d(GB)}{dG}\right)^2$$

so that
$$E\left(\frac{\partial^2 L}{\partial\alpha^2}\right)_{\alpha=0} = -\int_0^1\left(\frac{d(FA)}{dF}\right)^2 dF\int_0^1\left(\frac{d(GB)}{dG}\right)^2 dG$$

and
$$\operatorname{var}\hat{\alpha} = \frac{1}{n\int_0^1\left(\frac{d(FA)}{dF}\right)^2 dF\int_0^1\left(\frac{d(GB)}{dG}\right)^2 dG}.$$

9. Asymptotic relative efficiencies

As has already been shown in §§ 5–8 the mean value of the Spearman's rank correlation coefficient, r_s, Kendall's coefficient, t, the product moment correlation coefficient, r, and the maximum likelihood estimate $\hat{\alpha}$ are:

$$E(r_s) = 12\alpha\int_0^1 FA\,dF\int_0^1 GB\,dG \quad \text{or more generally} \quad 12\alpha\sum_i a_i\int_0^1 FA_i\,dF\int_0^1 GB_i\,dG;$$

$$E(t) = 8\alpha\int_0^1 FA\,dF\int_0^1 GB\,dG \quad \text{or more generally} \quad 8\alpha\sum_i a_i\int_0^1 FA_i\,dF\int_0^1 GB_i\,dG;$$

$$E(r) = \frac{\alpha}{\sigma_x\sigma_y}\int_0^1 X\frac{d(FA)}{dF}dF\int_0^1 Y\frac{d(GB)}{dG}dG \quad \text{or} \quad \frac{\alpha}{\sigma_x\sigma_y}\sum_i a_i\int_0^1 X\frac{d(FA)_i}{dF}dF\int_0^1 Y\frac{d(GB)_i}{dG}dG;$$

$$E(\hat{\alpha}) = \alpha.$$

The variances of r_s, t, r and $\hat{\alpha}$ under the hypothesis $\alpha = 0$ are:

$$\operatorname{var}(r_s) = \frac{1}{n-1}; \quad \operatorname{var}(t) = \frac{4}{9n}; \quad \operatorname{var}(r) = \frac{1}{n};$$

$$\operatorname{var}(\hat{\alpha}) = \frac{1}{n\int_0^1\left(\frac{d(FA)}{dF}\right)^2 dF\int_0^1\left(\frac{d(GB)}{dG}\right)^2 dG}$$

considering only the sample bivariate function without summation, the asymptotic relative efficiencies relative to the maximum likelihood estimate are, for r_s and t,

$$\frac{144\left(\int_0^1 FA\,dF\int_0^1 GB\,dG\right)^2}{\int_0^1\left(\frac{d(FA)}{dF}\right)^2 dF\int_0^1\left(\frac{d(GB)}{dG}\right)^2 dG}$$

and for r

$$\frac{\left(\int_0^1 X\frac{d(FA)}{dF}\,dF\int_0^1 Y\frac{d(GB)}{dG}\,dG\right)^2}{\sigma_x^2\sigma_y^2\int_0^1\left(\frac{d(FA)}{dF}\right)^2 dF\int_0^1\left(\frac{d(GB)}{dG}\right)^2 dG}.$$

It can be seen that each efficiency is the product of an element dependent only on functions associated with x and a second dependent on the corresponding functions of y. If we wish to find the functions of x and y that make these asymptotic relative efficiencies the greatest for each of the estimators of α, it suffices to maximize each of these elements separately. This property does not hold for the more general case and hence we will restrict attention to the simple bivariate function. For this class of bivariate distribution functions the two rank correlation estimators are equivalent in their asymptotic relative efficiences.

10. Determination of the 'best' bivariate distribution functions

(*a*) What are the functions $G(y)$ and $B(G(y))$, $F(x)$ and $A(F(x))$, that maximize the relative efficiencies of the rank correlation estimators when compared with the maximum likelihood estimators?

This problem can be solved by the Calculus of Variations. For this, we look for the function $A(F(x))$ that maximizes the variance of the maximum likelihood estimate subject to a fixed variance for the rank correlation coefficients.

To find the stationary value of $\int_0^1\left(\frac{d(FA)}{dF}\right)^2 dF$ subject to $\int_0^1 FA\,dF = \text{constant} = 1$, (say), put

$$\int_0^1\{(y')^2+\lambda y\}\,dx = V,$$

where $y = FA$ and $x = F$.

The function y that yields a stationary value for the integral is given by

$$0 = \frac{\partial V}{\partial y} - \frac{d}{dx}\left(\frac{\partial V}{\partial y'}\right),$$

$$0 = \lambda - 2y'',$$

or

$$y = \frac{\lambda x^2}{4} + cx + d.$$

Now $y = 0$ when $x = 0$ and $x = 1$; therefore

$$y = \frac{\lambda}{4}x(1-x) \quad \text{and} \quad A = \frac{\lambda}{4}(1-F).$$

The asymptotic relative efficiency for this choice of A is

$$144\left\{\int_0^1 x(1-x)\,dx\right\}^4\frac{\lambda_1^2\lambda_2^2}{16}\frac{16}{\lambda_1^2\lambda_2^2\left\{\int_0^1(1-2x)^2\,dx\right\}^2} = 1.$$

This incidentally proves that the stationary value we have found is the one corresponding to the greatest efficiency for rank correlation methods.

Referring back to § 2, this is the distribution quoted there as having a linear regression of the expectation of x given y when y is rectangularly distributed, and is one for which ranks and variate values are asymptotically perfectly correlated (Stuart, 1954).

(*b*) The determination of functions $G(y)$ and $B(G(y))$, $F(x)$ and $A(F(x))$, that maximize the relative efficiency of the product moment correlation estimator when compared with the maximum likelihood estimator requires that we find the stationary value of

$$\int_0^1 \left(\frac{d(FA)}{dF}\right)^2 dF \quad \text{subject to} \quad \int_0^1 x\frac{d(FA)}{dF}dF = \text{constant},$$

and x having mean μ and variance σ^2. We can, without loss of generality, choose the scale of x such that $\mu = 0$, $\sigma^2 = 1$. Put $\{d(FA)\}/dF = z$, $x = y$, $dF = dx$ and

$$V = \int_0^1 \{z^2 + \lambda_1 yz + \lambda_2 y + \lambda_3 y^2\}dx.$$

The required functions can be found as the solutions of

$$\lambda_1 z + \lambda_2 + 2\lambda_3 y = 0,$$
$$2z + \lambda_1 y = 0;$$

multiply each of the determining equations by y and integrate over x from 0 to 1, and use the relations

$$\int_0^1 yz\,dx = a, \quad \int_0^1 y\,dx = 0, \quad \int_0^1 y^2 dx = 1;$$

then

$$\lambda_1 a + 2\lambda_3 = 0 \quad \text{and} \quad 2a + \lambda_1 1 = 0.$$

Solving these equations and making the appropriate substitutions, including a return to the original variables, gives $\{d(FA)\}/dF = ax$ and thus

$$FA = a\int_0^F x\,dF$$

by the requirement, $FA = 0$ when $F = 0$.

The relationship between F and x is undetermined.

The asymptotic relative efficiency is

$$\left(\int_0^1 ax^2 dF\right)^4 \Big/ \left(\int_0^1 a^2x^2 dF\right)^2 = 1.$$

This again shows that this stationary value is indeed that which gives the greatest efficiency for the product moment correlation coefficient and not some least value.

(*c*) The stationary value for the asymptotic relative efficiencies, when comparing the product moment correlation coefficient and the rank correlation coefficients, is found to be neither a maximum nor a minimum value.

The position found is that of the intersection of the best forms for the rank correlation coefficients and the product moment correlation coefficient, namely,

$$z = FA = aF(1-F),$$
$$y = 3(1-2x) = 3(1-2F),$$
$$\int_0^x y\,dx = 3F(1-F) = z \quad \text{if} \quad a = +3.$$

Thus along line $A = 1-F$, the rank correlation methods are better than the product-moment correlation coefficient; along $A = \int_0^x x\,dF \Big/ \int_0^x dF$ the product-moment correlation coefficient is better than the rank correlation estimators. At the intersection, each is equally efficient with the maximum likelihood estimator.

11. SUMMARY

Summarizing these results we have:

(1) The product-moment correlation coefficient is fully efficient whenever the disturbing functions $A(x)$ and $B(y)$ are of the form

$$A(x) = \int_{-\infty}^{x} x\,dF \Big/ \int_{-\infty}^{x} dF \quad \text{and} \quad B(y) = \int_{-\infty}^{y} y\,dG \Big/ \int_{-\infty}^{y} dG,$$

where the origin of x is taken to be at the mean value of x, and that of y at its mean value.

(2) The two rank correlation coefficients considered are equally asymptotically efficient and are fully efficient if the disturbing functions $A(x)$ and $B(y)$ are given by

$$A(x) = 1-F(x), \quad B(y) = 1-G(y).$$

Recalling the result of 4·1 (*c*) we see that the product-moment correlation coefficient will be most efficient for those distributions for which the regression of y on x, and x on y are both linear. If we consider Spearman's rank correlation coefficient it is clear that here also we have a linear regression of $F(x)$ on $G(y)$ and $G(y)$ on $F(x)$ and the rank correlation coefficient corresponds to a product-moment correlation coefficient using expected scores, namely $i/(n+1)$ for the variates.

Hence, if a transformation of the x and y values can be found for which the regressions of the variables on each other are linear, the product-moment correlation coefficient in these transformed variables will be a fully efficient test for the association of the two variates. This can be achieved if we write $\{d(GB)\}/dG = \phi(y)$ and use the variable $\phi(y) = z$ as the transformed variable for y and the corresponding $w = \{d(FA)\}/dF$ as the transformed variable for x.

The product-moment correlation coefficient in these new variables will provide an asymptotically efficient test of association.

The transformation need not be monotone in character since,with the present class of bivariate distribution, the conditional distribution of x is affected by y only in the form $\{d(GB)\}/dG$; hence for the values of y for which $\{d(GB)\}/dG$ is the same it is immaterial which of the y values has in fact been observed and no information is lost in pooling them into a common value z.

The rank correlation corresponding to this statistic will be the product-moment correlation coefficient using the expected values of the order statistics in place of the observed values and would correspond therefore to scores

$$\int_0^1 F^{i-1} \frac{d(FA)}{dF} (1-F)^{N-1}\,dF$$

when $\{d(FA)\}/dF$ is a monotone function.

It is interesting to seek intuitive reasons why these last-mentioned statistics should be efficient. A property of the bivariate distribution considered is that the conditional dis-

tribution function of X, given the value of Y, is distorted by a function that is constant in shape for all y but of varying magnitude. As a consequence of this, it follows that there will be some values of X, independent of the Y chosen, for which the conditional probability density function shows the greatest deviation from the p.d.f. in the case of independence. We may expect that an efficient test will give greater weight to observations falling at these points and, since the function has some symmetry in the X and Y variables, the same will hold for the y values.

There will be some optimal relative weighting between the various x values that is independent of the chosen y and vice-versa. If each observation (x, y) is given a score $W(x, y)$ this implies

$$W(x, y|y) = w_1(x) \times \text{function of } y,$$

$$W(x, y|x) = w_2(y) \times \text{function of } x$$

and thus

$$W(x, y) = w_1(x)\, w_2(y) \times \text{constant}.$$

The test, therefore, seems likely to be a kind of product-moment correlation coefficient with x replaced by $w_1(x)$ and y replaced by $w_2(y)$. The analysis of this paper shows that

$$w_1(x) = \{d(FA)\}/dF \quad \text{and that} \quad w_2(y) = \{d(GB)\}/dG.$$

In a further paper it is hoped to extend the results of this paper to the general class of rank correlation coefficients introduced by Daniels (1944).

The author would like to express his thanks to Dr D. R. Cox for helpful criticism and advice during the preparation of this paper. This work was carried out as part of a course of postgraduate study in statistics at Birkbeck College, London.

References

Daniels, H. E. (1944). *Biometrika*, **33**, 129–35.
Daniels, H. E. (1950). *J. R. Statist. Soc.* B, **12**, 171–81.
Fieller, E. C., Hartley, H. O. & Pearson, E. S. (1957). *Biometrika*, **44**, 470–81.
Gumbel, E. J. (1958). *Comptes Rendus*, **246**, no. 19, 2717–19.
Hoeffding, W. (1947). *Biometrika*, **34**, 183–96.
Kendall, M. G. (1943). *Advanced Theory of Statistics*, **1**. London, Charles Griffin and Co. Ltd.
Konijn, H. S. (1956). *Ann. Math. Statist.* **27**, 300–23.
Morgenstern, D. (1956). *Mitt. Math. Statist.* **8**, 234–5.
Pitman, E. J. G. (1948). Unpublished lecture notes. University of N. Carolina.
Stuart, A. (1954). *Brit. J. Psychol. Statist. Sect.* **7**, 37–44.

Biometrika (1960), **47**, 3 *and* 4, *p.* 325
Printed in Great Britain

Two expansions for the quadrivariate normal integral*

BY J. A. McFADDEN
Purdue University

1. INTRODUCTION

Let X_1, X_2, X_3 and X_4 obey a quadrivariate normal distribution with zero means, and let P_4 be the value of the quadrivariate normal integral, i.e. the probability that X_1, X_2, X_3 and X_4 are simultaneously positive. The generalized tetrachoric series for P_4, as given by Aitken (unpublished), Kendall (1941, 1945) (see also Kendall & Stuart, 1958, pp. 350–4), and Moran (1948), are not well suited for computation. For the case in which all six correlation coefficients are equal, the series has been summed approximately by McFadden (1956).

For special numerical values of the correlation matrix, exact results for P_4 have been given by Schläfli (1858, 1860), Anis & Lloyd (1953), and Plackett (1954). Methods for numerical integration in more general cases have been given by Plackett (1954), Ihm (1959), and John (1959). Numerical methods for integration when all the correlation coefficients are equal have been provided by Ruben (1954) and Moran (1956).

In this paper we present two series expansions for P_4 which are well suited for computation. The first case occurs when X_1, X_2, X_3 and X_4 are successive measurements from a stationary Gaussian Markov process (with zero mean), or, equivalently, when the inverse of the correlation matrix has zero elements except on the main diagonal and immediately adjacent to it. The second case occurs when the correlation matrix itself has zero elements except on the main diagonal and adjacent to it. We shall then show that these two cases are related by a simple transformation.

In a previous paper, McFadden (1955, eq. (39)) has expressed P_4 in terms of the various product moments of the set of random variables Y_i which assume the sign of the normal variables X_i. Let

$$\left.\begin{aligned} Y_i &= 1 \quad \text{when} \quad X_i \geqslant 0 \\ &= -1 \quad \text{when} \quad X_i < 0 \end{aligned}\right\} \quad (i = 1, 2, 3, 4). \tag{1}$$

Then

$$P_4 = \frac{1}{16}\left[1 + \sum_{j>i\geqslant 1}^{4} E(Y_iY_j) + E(Y_1Y_2Y_3Y_4)\right], \tag{2}$$

where

$$E(Y_iY_j) = \frac{2}{\pi}\sin^{-1}\rho_{ij}, \tag{3}$$

and ρ_{ij} is the correlation coefficient between X_i and X_j. In general, the fourth product moment in (2) depends on all six correlation coefficients, but in the two cases to be considered it depends on only three. In each of the two cases we shall derive a series expansion for $E(Y_1Y_2Y_3Y_4)$ in powers of ρ_{23}.

* This work was supported by the Office of Naval Research, U.S. Navy. Some of these results were first presented to the Institute of Mathematical Statistics at Lafayette, Indiana, 9 April 1960. (McFadden 1960.)

2. The Markov case

Suppose that X_1, X_2, X_3 and X_4 are successive measurements from a stationary Gaussian Markov process with zero mean. For convenience we shall take the variance of X_i to be unity. Then it is known (Doob, 1942, p. 354) that the autocorrelation function is exponential, and consequently

$$\rho_{13} = \rho_{12}\rho_{23}; \quad \rho_{14} = \rho_{12}\rho_{23}\rho_{34}; \quad \rho_{24} = \rho_{23}\rho_{34}. \tag{4}$$

Thus we may express the six second-order moments in (2) in terms of ρ_{12}, ρ_{23} and ρ_{34}.

Let us now derive the fourth-order moment $E(Y_1Y_2Y_3Y_4)$. For the Markov case, we define this function as follows

$$E(Y_1Y_2Y_3Y_4) \equiv W(\rho_{12}, \rho_{23}, \rho_{34}). \tag{5}$$

Because of the Markov property, we may express the joint probability density of the X_i as follows (Doob, 1942, p. 353)

$$f(x_1, x_2, x_3, x_4) = f_{12}(x_1, x_2) f_{34}(x_3, x_4) \frac{f_{23}(x_2, x_3)}{f_2(x_2) f_3(x_3)}, \tag{6}$$

where $f_i(x_i)$ is the marginal density function for the standardized normal distribution and $f_{ij}(x_i, x_j)$ is the bivariate density function. Thus

$$f_i(x_i) = (2\pi)^{-\frac{1}{2}} \exp(-\tfrac{1}{2}x_i^2); \tag{7}$$

$$f_{ij}(x_i, x_j) = (2\pi)^{-1}(1-\rho_{ij}^2)^{-\frac{1}{2}} \exp[-\tfrac{1}{2}(1-\rho_{ij}^2)^{-1}(x_i^2 - 2\rho_{ij}x_ix_j + x_j^2)]. \tag{8}$$

[The resulting expression for P_4 will actually hold for any quadrivariate normal density of the type (6) in which the four variances are arbitrary and not necessarily equal; we have taken them equal to unity only for convenience. The assumption of stationarity requires that the variances be equal, but the assumption of a stationary Markov process could be replaced by the hypothesis (6) with arbitrary variances. Note also that the products x_1x_3, x_1x_4 and x_2x_4 do not appear.]

We may expand the last factor of (6) in Mehler's series of Hermite polynomials (Cramér, 1946, p. 133).

$$\frac{f_{23}(x_2, x_3)}{f_2(x_2) f_3(x_3)} \equiv (1-\rho_{23}^2)^{-\frac{1}{2}} \exp[-\tfrac{1}{2}(1-\rho_{23}^2)^{-1}(\rho_{23}^2x_2^2 - 2\rho_{23}x_2x_3 + \rho_{23}^2x_3^2)] = \sum_{\nu=0}^{\infty} \frac{H_\nu(x_2) H_\nu(x_3)}{\nu!} \rho_{23}^\nu. \tag{9}$$

Thus we may expand (6) in a series, each term containing two factors, one depending on x_1 and x_2, and the other depending on x_3 and x_4. By the definition (5), we find that

$$W(\rho_{12}, \rho_{23}, \rho_{34}) = \sum_{\nu=0}^{\infty} \frac{\rho_{23}^\nu}{\nu!} \alpha_\nu(\rho_{12}) \alpha_\nu(\rho_{34}), \tag{10}$$

where the expectations

$$\left.\begin{aligned} \alpha_\nu(\rho_{12}) &= E[Y_1Y_2H_\nu(X_2)], \\ \alpha_\nu(\rho_{34}) &= E[Y_3Y_4H_\nu(X_3)] \end{aligned}\right\} \tag{11}$$

require integration in only two dimensions instead of four.

By the symmetry of the bivariate normal distribution, we find that these functions α_ν vanish when ν is odd. In order to obtain the even cases, we must first calculate the general moment $E(Y_1Y_2X_2^{2n})$, where n is a non-negative integer. The first few cases may be obtained from the results of Kamat (1953), but the following method yields the general result.

Rice (1944–45, § 3·5) has given the formula

$$\int_0^\infty \int_0^\infty x^l y^m \exp(-x^2-y^2-2axy)\,dx\,dy = \frac{1}{4}\sum_{r=0}^{\infty} \frac{(-2a)^r}{r!}\,\Gamma\left(\frac{l+r+1}{2}\right)\Gamma\left(\frac{m+r+1}{2}\right). \quad (12)$$

With $l=0$, $m=2n$ and $a=-\rho_{12}$, a change of variables in (12) yields the contribution of the first quadrant to the desired moment. The contribution of the third quadrant is identical. The contributions of the second and fourth quadrants are also identical to that of the first, except that the expression is negative and furthermore ρ_{12} must be replaced by $-\rho_{12}$. As a result, all the terms of (12) cancel for which r is even. With $r=2s+1$, we find that

$$\begin{aligned} E(Y_1Y_2X_2^{2n}) &= \frac{2^n}{\pi}(1-\rho_{12}^2)^{n+\frac{1}{2}} \sum_{s=0}^{\infty} \frac{(2\rho_{12})^{2s+1}}{(2s+1)!}\,\Gamma(s+1)\,\Gamma(n+s+1) \\ &= \frac{2^{n+1}}{\pi}\,n!\,\rho_{12}(1-\rho_{12}^2)^{n+\frac{1}{2}}\,F(1,n+1;\tfrac{3}{2};\rho_{12}^2) \\ &= \frac{2^{n+1}}{\pi}\,n!\,\rho_{12}F(\tfrac{1}{2},\tfrac{1}{2}-n;\tfrac{3}{2};\rho_{12}^2). \end{aligned} \quad (13)$$

This last step requires Kummer's transformation (Erdélyi *et al.* 1953, p. 105, eq. (2)).

Let us now express the Hermite polynomials of even order as follows (Magnus & Oberhettinger, 1954, p. 82):

$$H_{2\mu}(X_2) = (-2)^\mu(\tfrac{1}{2})_\mu \sum_{n=0}^{\infty} \frac{(-\mu)_n}{(\frac{1}{2})_n\,n!}\left(\frac{X_2^2}{2}\right)^n, \quad (14)$$

where $(a)_n = \Gamma(a+n)/\Gamma(a)$. If we substitute (14) into (11), interchange the order of the operations, and employ (13), we find that

$$\begin{aligned} \alpha_{2\mu}(\rho_{12}) &= (-2)^\mu(\tfrac{1}{2})_\mu \sum_{n=0}^{\infty} \frac{(-\mu)_n}{2^n(\frac{1}{2})_n\,n!}\,E(Y_1Y_2X_2^{2n}) \\ &= \frac{2}{\pi}(-2)^\mu(\tfrac{1}{2})_\mu\rho_{12} \sum_{n=0}^{\infty} \frac{(-\mu)_n}{(\frac{1}{2})_n} \sum_{s=0}^{\infty} \frac{(\frac{1}{2})_s(\frac{1}{2}-n)_s}{(\frac{3}{2})_s}\,\frac{\rho_{12}^{2s}}{s!}. \end{aligned} \quad (15)$$

We now interchange the order of the summations and use the identity

$$(\tfrac{1}{2}-n)_s = \frac{(\frac{1}{2})_s(\frac{1}{2})_n}{(\frac{1}{2}-s)_n}; \quad (16)$$

then we find that (15) becomes the following

$$\alpha_{2\mu}(\rho_{12}) = \frac{2}{\pi}(-2)^\mu(\tfrac{1}{2})_\mu\rho_{12} \sum_{s=0}^{\infty} \frac{(\frac{1}{2})_s(\frac{1}{2})_s}{(\frac{3}{2})_s}\,\frac{\rho_{12}^{2s}}{s!}\,F(1,-\mu;\tfrac{1}{2}-s;1). \quad (17)$$

We now make use of the formula for the hypergeometric function of unit argument (Erdélyi *et al.* 1953, p. 61, eq. (14)). Ordinarily we could employ this formula only when $\mu > s+\frac{1}{2}$, but because the series terminates we may effect the usual derivation (loc. cit.) without this restriction. After simplifying the expression (17), we find that

$$\alpha_{2\mu}(\rho_{12}) = \frac{2}{\pi}(-2)^\mu(-\tfrac{1}{2})_\mu\rho_{12}F(\tfrac{1}{2},\tfrac{1}{2}-\mu;\tfrac{3}{2}-\mu;\rho_{12}^2) \quad (18)$$

$$= \frac{2}{\pi}(-2)^\mu(-\tfrac{1}{2})_\mu\rho_{12}(1-\rho_{12}^2)^{\frac{1}{2}}F(1,1-\mu;\tfrac{3}{2}-\mu;\rho_{12}^2), \quad (19)$$

the last step being Kummer's transformation.

When $\mu = 0$, (18) yields an arcsine function (Erdélyi *et al.* 1953, p. 102).

$$\alpha_0(\rho_{12}) = \frac{2}{\pi}\sin^{-1}\rho_{12}. \tag{20}$$

When $\mu \geqslant 1$ the hypergeometric series in (19) become polynomials of degree $(\mu - 1)$ in ρ_{12}^2. Thus the terms of the series (10) for $W(\rho_{12}, \rho_{23}, \rho_{34})$ are well suited for computation. As we shall prove in the appendix, (10) converges for all values in the range $0 \leqslant \rho_{12}^2 \leqslant 1$, $0 \leqslant \rho_{23}^2 \leqslant 1$, $0 \leqslant \rho_{34}^2 \leqslant 1$.

The first four terms of (10) appear as follows

$$W(\rho_{12}, \rho_{23}, \rho_{34}) = \frac{4}{\pi^2}\left\{\sin^{-1}\rho_{12}\sin^{-1}\rho_{34} + \rho_{12}\rho_{34}(1-\rho_{12}^2)^{\frac{1}{2}}(1-\rho_{34}^2)^{\frac{1}{2}} \times \left[\frac{\rho_{23}^2}{2!} + \frac{\rho_{23}^4}{4!}(1+2\rho_{12}^2)(1+2\rho_{34}^2) + \frac{\rho_{23}^6}{6!}(3+4\rho_{12}^2+8\rho_{12}^4)(3+4\rho_{34}^2+8\rho_{34}^4) + \ldots\right]\right\}. \tag{21}$$

We may test the validity of (10) in several examples. First, let us consider degenerate cases. If $\rho_{12} = 0$, then X_1 is independent of X_2, X_3 and X_4, and $E(Y_1Y_2Y_3Y_4)$ becomes $E(Y_1)E(Y_2Y_3Y_4)$, which is zero. The series (10) also vanishes.

If $\rho_{23} = 0$, then X_1 and X_2 are independent of X_3 and X_4, and $E(Y_1Y_2Y_3Y_4)$ becomes $E(Y_1Y_2)E(Y_3Y_4)$, or, by (3), $(4/\pi^2)\sin^{-1}\rho_{12}\sin^{-1}\rho_{34}$. This result agrees with the first term of (10); the other terms vanish.

If $\rho_{12} = 1$, then $Y_1Y_2 = Y_1^2 = 1$ and $E(Y_1Y_2Y_3Y_4) = E(Y_3Y_4) = (2/\pi)\sin^{-1}\rho_{34}$. This result agrees with the first term of (10); the other terms vanish.

If $\rho_{23} = 1$, then $Y_2Y_3 = 1$ and, by (3) and (4), $E(Y_1Y_2Y_3Y_4) = E(Y_1Y_4) = (2/\pi)\sin^{-1}(\rho_{12}\rho_{34})$. It is not obvious that (10) reduces to the same result, but the identity can be proved.

One non-degenerate example is provided by the exact result of Anis & Lloyd (1953), who have shown, in effect, that when $\rho_{12} = \rho_{34} = \sqrt{(6)}/4$, $\rho_{13} = \rho_{24} = \sqrt{(6)}/6$, $\rho_{14} = \frac{1}{4}$, and $\rho_{23} = \frac{2}{3}$, then $P_4 = \frac{1}{5}$. These correlation coefficients satisfy (4). We may obtain the true value of W by (2), and compare it with the result of a finite number of terms of (21).

A more extreme test of the usefulness of (21) is provided by the results of Rice (1958, p. 623). When $\rho_{12} = \rho_{23} = \rho_{34}$ in the Markov case, Rice has calculated $2P_4$ by numerical integration for several large values of the correlation coefficients. We can compare the results as before.

We have made the comparison in W rather than in P_4 so as to accentuate the problem. It is clear that, if the second-order moments in (2) are positive, then the percentage error will be smaller in P_4 than in W.

In the last column of the following table, we have used a non-linear sequence-to-sequence transformation due to Shanks (1955). If W_n is the sum of n terms of (21), we apply Shanks's first-order transformation to the quantities 0, $W_3 - W_2$, and $W_4 - W_2$ and add the result to W_2. This result we shall call $e_1(W_3)$.

			True	True	From (21)	
ρ_{12}	ρ_{23}	ρ_{34}	P_4	W	W_4	$e_1(W_3)$
$\sqrt{(6)}/4$	2/3	$\sqrt{(6)}/4$	1/5	1/5	0·19990	0·19998
0·9835	0·9835	0·9835	0·868/2	0·82	0·7956	0·8069
0·9672	0·9672	0·9672	0·814/2	0·74	0·7238	0·7367
0·9512	0·9512	0·9512	0·774/2	0·69	0·6719	0·6842
0·9355	0·9355	0·9355	0·742/2	0·66	0·6297	0·6401
0·8752	0·8752	0·8752	0·641/2	0·51	0·5048	0·5092

We may conclude from the table that four terms of the series (21) provide an adequate approximation to the fourth product moment W (and, when the correlation coefficients are positive, an excellent approximation to the quadrivariate normal integral P_4), even when the correlation coefficients are relatively large.

One application in which the series (10) leads to an exact result is to a problem involving noise in an electric circuit. If a stationary Gaussian Markov process (with zero mean) is infinitely clipped at the levels ± 1 as in (1), let the resulting binary process be the input to an *RC* low-pass filter. The fourth moment of the output of this filter is a fourfold integral involving $W(\rho_{12}, \rho_{23}, \rho_{34})$. Using (10), we can integrate term by term and express the resulting series in closed form (McFadden, unpublished).

3. The case in which $\rho_{13} = \rho_{14} = \rho_{24} = 0$

Suppose again that X_1, X_2, X_3 and X_4 obey a quadrivariate normal distribution with zero means and arbitrary variances. Instead of assuming the relation (4), let us assume that

$$\rho_{13} = \rho_{14} = \rho_{24} = 0. \tag{22}$$

These situations have been studied by Schläfli (1858, 1860) and van der Vaart (1955). (Schläfli used the term 'orthoschemon' to describe this case.) In order that the determinant of the correlation matrix be positive, we must prescribe that

$$\rho_{23}^2 < (1-\rho_{12}^2)(1-\rho_{34}^2). \tag{23}$$

For this case we define the fourth product moment by a different symbol:

$$E(Y_1Y_2Y_3Y_4) \equiv \mathscr{W}(\rho_{12}, \rho_{23}, \rho_{34}). \tag{24}$$

Let us now employ the generalized tetrachoric series (Moran, 1948). After a sorting of terms, we can manipulate Moran's series for the general correlation matrix into the form (2), where

$$\begin{aligned}(\tfrac{1}{4}\pi^2)\,E(Y_1Y_2Y_3Y_4) = {}& \rho_{12}\rho_{34}\,\mathscr{F}(\tfrac{1}{2},\tfrac{1}{2},\tfrac{1}{2},\tfrac{1}{2};\tfrac{3}{2},\tfrac{1}{2},\tfrac{1}{2},\tfrac{1}{2},\tfrac{1}{2},\tfrac{3}{2};\rho_{12}^2,\rho_{13}^2,\rho_{14}^2,\rho_{23}^2,\rho_{24}^2,\rho_{34}^2)\\ &+\rho_{13}\rho_{24}\,\mathscr{F}(\tfrac{1}{2},\tfrac{1}{2},\tfrac{1}{2},\tfrac{1}{2};\tfrac{1}{2},\tfrac{3}{2},\tfrac{1}{2},\tfrac{1}{2},\tfrac{3}{2},\tfrac{1}{2};\rho_{12}^2,\rho_{13}^2,\rho_{14}^2,\rho_{23}^2,\rho_{24}^2,\rho_{34}^2)\\ &+\rho_{14}\rho_{23}\,\mathscr{F}(\tfrac{1}{2},\tfrac{1}{2},\tfrac{1}{2},\tfrac{1}{2};\tfrac{1}{2},\tfrac{1}{2},\tfrac{3}{2},\tfrac{3}{2},\tfrac{1}{2},\tfrac{1}{2};\rho_{12}^2,\rho_{13}^2,\rho_{14}^2,\rho_{23}^2,\rho_{24}^2,\rho_{34}^2)\\ &-\rho_{12}\rho_{13}\rho_{14}\,\mathscr{F}(\tfrac{3}{2},\tfrac{1}{2},\tfrac{1}{2},\tfrac{1}{2};\tfrac{3}{2},\tfrac{3}{2},\tfrac{3}{2},\tfrac{1}{2},\tfrac{1}{2},\tfrac{1}{2};\rho_{12}^2,\rho_{13}^2,\rho_{14}^2,\rho_{23}^2,\rho_{24}^2,\rho_{34}^2)\\ &-\rho_{12}\rho_{23}\rho_{24}\,\mathscr{F}(\tfrac{1}{2},\tfrac{3}{2},\tfrac{1}{2},\tfrac{1}{2};\tfrac{3}{2},\tfrac{1}{2},\tfrac{1}{2},\tfrac{3}{2},\tfrac{3}{2},\tfrac{1}{2};\rho_{12}^2,\rho_{13}^2,\rho_{14}^2,\rho_{23}^2,\rho_{24}^2,\rho_{34}^2)\\ &-\rho_{13}\rho_{23}\rho_{34}\,\mathscr{F}(\tfrac{1}{2},\tfrac{1}{2},\tfrac{3}{2},\tfrac{1}{2};\tfrac{1}{2},\tfrac{3}{2},\tfrac{1}{2},\tfrac{3}{2},\tfrac{1}{2},\tfrac{3}{2};\rho_{12}^2,\rho_{13}^2,\rho_{14}^2,\rho_{23}^2,\rho_{24}^2,\rho_{34}^2)\\ &-\rho_{14}\rho_{24}\rho_{34}\,\mathscr{F}(\tfrac{1}{2},\tfrac{1}{2},\tfrac{1}{2},\tfrac{3}{2};\tfrac{1}{2},\tfrac{1}{2},\tfrac{3}{2},\tfrac{1}{2},\tfrac{3}{2},\tfrac{3}{2};\rho_{12}^2,\rho_{13}^2,\rho_{14}^2,\rho_{23}^2,\rho_{24}^2,\rho_{34}^2)\\ &+\rho_{12}\rho_{13}\rho_{14}\rho_{23}\rho_{24}\rho_{34}\,\mathscr{F}(\tfrac{3}{2},\tfrac{3}{2},\tfrac{3}{2},\tfrac{3}{2};\tfrac{3}{2},\tfrac{3}{2},\tfrac{3}{2},\tfrac{3}{2},\tfrac{3}{2},\tfrac{3}{2};\rho_{12}^2,\rho_{13}^2,\rho_{14}^2,\rho_{23}^2,\rho_{24}^2,\rho_{34}^2),\end{aligned} \tag{25}$$

where $\mathscr{F}(\ )$ is a generalized hypergeometric series, defined as follows:

$$\begin{aligned}&\mathscr{F}(a_1,a_2,a_3,a_4;b_1,b_2,b_3,b_4,b_5,b_6;x_1,x_2,x_3,x_4,x_5,x_6)\\ &\quad\equiv \sum_{l,m,n,p,q,r=0}^{\infty} \frac{(a_1)_{l+m+n}(a_2)_{l+p+q}(a_3)_{m+p+r}(a_4)_{n+q+r}}{(b_1)_l(b_2)_m(b_3)_n(b_4)_p(b_5)_q(b_6)_r}\,\frac{x_1^l x_2^m x_3^n x_4^p x_5^q x_6^r}{l!\,m!\,n!\,p!\,q!\,r!}.\end{aligned} \tag{26}$$

For the particular case (22), all but one of the series in (25) vanish and $\mathscr{W}$ becomes the triple series,

$$\mathscr{W}(\rho_{12},\rho_{23},\rho_{34}) = \frac{4}{\pi^2}\rho_{12}\rho_{34}\sum_{l,p,r=0}^{\infty}\frac{(\frac{1}{2})_l(\frac{1}{2})_{l+p}(\frac{1}{2})_{p+r}(\frac{1}{2})_r}{(\frac{3}{2})_l(\frac{1}{2})_p(\frac{3}{2})_r}\,\frac{\rho_{12}^{2l}\rho_{23}^{2p}\rho_{34}^{2r}}{l!\,p!\,r!}. \tag{27}$$

Now
$$(\tfrac{1}{2})_{l+p} = (\tfrac{1}{2}+p)_l (\tfrac{1}{2})_p; \tag{28}$$
therefore we may rewrite (27) as follows
$$\mathscr{W}(\rho_{12}, \rho_{23}, \rho_{34}) = \sum_{p=0}^{\infty} \frac{(\frac{1}{2})_p}{p!} \rho_{23}^{2p} S_p(\rho_{12}) S_p(\rho_{34}), \tag{29}$$
where
$$S_p(\rho_{12}) = \frac{2}{\pi} \rho_{12} F(\tfrac{1}{2}, \tfrac{1}{2}+p; \tfrac{3}{2}; \rho_{12}^2) \tag{30}$$
$$= \frac{2}{\pi} \rho_{12} (1-\rho_{12}^2)^{\frac{1}{2}-p} F(1, 1-p; \tfrac{3}{2}; \rho_{12}^2), \tag{31}$$
the last step being Kummer's transformation.

As before, when $p = 0$, (30) yields an arcsine function, and when $p \geqslant 1$ the series in (31) become polynomials of degree $(p-1)$ in ρ_{12}^2. Thus these terms, too, are well suited for computation. As we shall prove in the appendix, (29) converges when $0 \leqslant \rho_{12}^2 < 1$, $0 \leqslant \rho_{34}^2 < 1$, and (23) is satisfied.

The first four terms of (29) appear as follows
$$\mathscr{W}(\rho_{12}, \rho_{23}, \rho_{34}) = \frac{4}{\pi^2} \left\{ \sin^{-1}\rho_{12} \sin^{-1}\rho_{34} + \frac{\rho_{12}\rho_{34}}{(1-\rho_{12}^2)^{\frac{1}{2}} (1-\rho_{34}^2)^{\frac{1}{2}}} \left[\frac{\rho_{23}^2}{2!} + \frac{\rho_{23}^4}{4!} \frac{(3-2\rho_{12}^2)(3-2\rho_{34}^2)}{(1-\rho_{12}^2)(1-\rho_{34}^2)} + \frac{\rho_{23}^6}{6!} \frac{(15-20\rho_{12}^2+8\rho_{12}^4)(15-20\rho_{34}^2+8\rho_{34}^4)}{(1-\rho_{12}^2)^2 (1-\rho_{34}^2)^2} + \ldots \right] \right\}. \tag{32}$$

If one of the correlation coefficients vanishes, $\mathscr{W}$ reduces to zero or to two arcsines as before.

In a large number of non-trivial examples of the case (22), Schläfli (1860, p. 98) has given exact values for P_4. [These results have been summarized by Plackett (1954, p. 359) and some of them also by Coxeter (1935).] As before, we may compute the exact value of $\mathscr{W}$ by (2), and compare the partial sums of (32), with and without the non-linear transform. Some results are given below.

			True	True	From (32)	
ρ_{12}	ρ_{23}	ρ_{34}	P_4	$\mathscr{W}$	$\mathscr{W}_4$	$e_1(\mathscr{W}_3)$
$-\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{2}$	$\frac{1}{120}$	$\frac{2}{15} \cong 0{\cdot}13333$	0·13293	0·13329
$-1/\sqrt{2}$	$-\frac{1}{2}$	$-\frac{1}{2}$	$\frac{1}{384}$	$\frac{5}{24} \cong 0{\cdot}20833$	0·20669	0·20799
$-\frac{1}{2}$	$-1/\sqrt{2}$	$-\frac{1}{2}$	$\frac{1}{1152}$	$\frac{13}{72} \cong 0{\cdot}18056$	0·16869	0·17764
$-\sin(\pi/10)$	$-\frac{1}{2}$	$-\frac{1}{2}$	$\frac{191}{14400}$	$\frac{71}{900} \cong 0{\cdot}078889$	0·078715	0·078873

[Incidentally, the first example can be done combinatorially, without assuming normality, in a manner similar to the method described by Plackett (1954, p. 360).]

Again we observe that the agreement is good, although poorer near the edge of the region of convergence, as in the third example.

A somewhat more complicated series for P_4 in the case (22) has been given by Coxeter (1935). [For the interpretation of Coxeter's notation in terms of the present problem, we refer the reader to Plackett (1954, p. 359).]

4. The relation between the two cases

A simple relation exists between the functions $W(\rho_{12}, \rho_{23}, \rho_{34})$ and $\mathscr{W}(\lambda_{12}, \lambda_{23}, \lambda_{34})$ of the two previous sections. To avoid confusion, we have introduced new variables λ_{ij} for the elements of the correlation matrix having the property (22).

As a special case of an identity for hypergeometric functions (Erdélyi *et al.* 1953, p. 106, eq. (26)), we can show that*

$$F(1, 1-\mu; \tfrac{3}{2}-\mu; x) = -\frac{(\frac{1}{2})_\mu}{(-\frac{1}{2})_\mu} F(1, 1-\mu; \tfrac{3}{2}; 1-x) \quad (\mu = 1, 2, 3, \ldots). \tag{33}$$

We then find from (19) and (31) that

$$\begin{aligned}\alpha_{2\mu}(\rho_{12}) &= -\frac{2}{\pi}(-2)^\mu (\tfrac{1}{2})_\mu \rho_{12}(1-\rho_{12}^2)^{\frac{1}{2}} F(1, 1-\mu; \tfrac{3}{2}; 1-\rho_{12}^2) \\ &= -(-2)^\mu (\tfrac{1}{2})_\mu \rho_{12}^{2\mu} S_\mu[(1-\rho_{12}^2)^{\frac{1}{2}}] \quad (\mu = 1, 2, 3, \ldots).\end{aligned} \tag{34}$$

We may now obtain the general even term in (10) ($\mu \neq 0$).

$$\frac{\rho_{23}^{2\mu}}{(2\mu)!}\alpha_{2\mu}(\rho_{12})\,\alpha_{2\mu}(\rho_{34}) = \frac{(\frac{1}{2})_\mu}{\mu!}(\rho_{12}\rho_{23}\rho_{34})^{2\mu} S_\mu[(1-\rho_{12}^2)^{\frac{1}{2}}] S_\mu[(1-\rho_{34}^2)^{\frac{1}{2}}]. \tag{35}$$

This is the general term of $\mathscr{W}(\lambda_{12}, \lambda_{23}, \lambda_{34})$ [cf. (29)] with λ_{12} replaced by $(1-\rho_{12}^2)^{\frac{1}{2}}$, λ_{34} by $(1-\rho_{34}^2)^{\frac{1}{2}}$, and λ_{23} by $\rho_{12}\rho_{23}\rho_{34}$ ($= \rho_{14}$). Then we may remove the zeroth terms of (10) and (29) and derive the identity

$$\mathscr{W}[(1-\rho_{12}^2)^{\frac{1}{2}}, \rho_{12}\rho_{23}\rho_{34}, (1-\rho_{34}^2)^{\frac{1}{2}}] = W(\rho_{12}, \rho_{23}, \rho_{34}) + 1 - \frac{2}{\pi}(\sin^{-1}\rho_{12} + \sin^{-1}\rho_{34}), \tag{36}$$

or, conversely,

$$\begin{aligned}W\left[(1-\lambda_{12}^2)^{\frac{1}{2}}, \frac{\lambda_{23}}{(1-\lambda_{12}^2)^{\frac{1}{2}}(1-\lambda_{34}^2)^{\frac{1}{2}}}, (1-\lambda_{34}^2)^{\frac{1}{2}}\right] \\ = \mathscr{W}(\lambda_{12}, \lambda_{23}, \lambda_{34}) + 1 - \frac{2}{\pi}(\sin^{-1}\lambda_{12} + \sin^{-1}\lambda_{34}).\end{aligned} \tag{37}$$

We may easily effect sign changes in the arguments above by considering that $W(\rho_{12}, \rho_{23}, \rho_{34})$ is an odd function of ρ_{12} and ρ_{34} and an even function of ρ_{23}, and similarly for $\mathscr{W}$.

Thus if we know P_4 for either of the cases in the previous sections, we can find a corresponding solution for the other case by means of these transformations. Evidently this transformation is a special case of a 'dissection of a plagioschemon [i.e. general correlation matrix] into orthoschemons' [i.e. all elements zero except on the main diagonal and adjacent to it], as considered by Schläfli (1858, p. 293). In this case the 'plagioschemon' is the Markov case (4).

As a special case of this transformation, consider the solution of Anis & Lloyd in § 2. If $\rho_{12} = \rho_{34} = \sqrt{(6)}/4$ and $\rho_{23} = \sqrt{(6)}/6$, then the arguments of $\mathscr{W}$ in (36) are $\lambda_{12} = \lambda_{34} = \sqrt{(10)}/4$ and $\lambda_{23} = \frac{1}{4}$. Then by (36),

$$\begin{aligned}\mathscr{W}(-\sqrt{(10)}/4, -\tfrac{1}{4}, -\sqrt{(10)}/4) &= \mathscr{W}(\sqrt{(10)}/4, \tfrac{1}{4}, \sqrt{(10)}/4) \\ &= \frac{6}{5} - \frac{4}{\pi}\sin^{-1}\frac{\sqrt{6}}{4}.\end{aligned} \tag{38}$$

Then by (2), the corresponding value of P_4 for these negative arguments is the following:

$$\begin{aligned}P_4 &= \frac{1}{16}\left[1 - \frac{4}{\pi}\sin^{-1}\frac{\sqrt{10}}{4} - \frac{2}{\pi}\sin^{-1}\frac{1}{4} + \frac{6}{5} - \frac{4}{\pi}\sin^{-1}\frac{\sqrt{6}}{4}\right] \\ &= \frac{1}{80} - \frac{1}{8\pi}\sin^{-1}\tfrac{1}{4}.\end{aligned} \tag{39}$$

* This result was first shown to the author by Prof. I. Marx.

This result agrees with one of Schläfli's examples [1860, p. 98, $f(\lambda, 2\lambda, \lambda)$] (also Plackett, 1954, p. 359, 13th item).

Conversely, from all of Schläfli's exact solutions we may derive corresponding solutions for the Markov case, thus adding to the catalogue of exact solutions for P_4.

5. Conclusion

In this paper we have derived two new expansions for the quadrivariate normal integral, the first being applicable in the stationary Markov case (4) and the second being applicable when the correlation coefficients are all zero except on the main diagonal and adjacent to it. These expansions are well suited for computation.

In addition, we have shown that either one of the two problems considered is reducible to the other by means of a simple transformation.

Appendix*

Let us now establish the convergence of the series (10) and (29), beginning with the latter. For large p, (31) becomes asymptotically (Erdélyi *et al.* 1953, p. 77, eq. (14))

$$S_p(\rho_{12}) \sim \frac{2}{\pi}\rho_{12}(1-\rho_{12}^2)^{\frac{1}{2}-p}\frac{1}{2(p-1)\rho_{12}^2}\left[1+O\left(\frac{1}{p\rho_{12}^2}\right)\right], \tag{40}$$

where $\rho_{12}^2 > 0$. Then

$$\lim_{p\to\infty}\frac{S_{p+1}(\rho_{12})}{S_p(\rho_{12})} = \frac{1}{1-\rho_{12}^2}, \tag{41}$$

and it follows directly from (31) that the limit is also valid at $\rho_{12} = 0$. Therefore, using the ratio test, we find that (29) converges everywhere in the range

$$\left.\begin{aligned} \rho_{23}^2 &< (1-\rho_{12}^2)(1-\rho_{34}^2), \\ 0 \leqslant \rho_{12}^2 &< 1, \\ 0 \leqslant \rho_{34}^2 &< 1, \end{aligned}\right\} \tag{42}$$

which is the region of importance in the present problem.

Consider now the other series (10). By (41), the limit of the ratio of successive values of (34) is, for $0 \leqslant \rho_{12}^2 \leqslant 1$,

$$\lim_{\mu\to\infty}\frac{\alpha_{2\mu+2}(\rho_{12})}{\alpha_{2\mu}(\rho_{12})} = 1. \tag{43}$$

Thus, by the ratio test, we find that (10) converges everywhere in the range,

$$\left.\begin{aligned} 0 \leqslant \rho_{12}^2 &\leqslant 1, \\ 0 \leqslant \rho_{23}^2 &< 1, \\ 0 \leqslant \rho_{34}^2 &\leqslant 1. \end{aligned}\right\} \tag{44}$$

We can show that the series also converges when $\rho_{23}^2 = 1$, as in § 2; therefore, by a property of power series, we may establish uniform convergence near $\rho_{23}^2 = 1$.

References

Anis, A. A. & Lloyd, E. H. (1953). On the range of partial sums of a finite number of independent normal variates. *Biometrika*, **40**, 35–42.

Coxeter, H. S. M. (1935). The functions of Schläfli and Lobatchefsky. *Quart. J. Math.* **6**, 13–29.

Cramér, H. (1946). *Mathematical Methods of Statistics*. Princeton University Press.

Doob, J. L. (1942). The Brownian movement and stochastic equations. *Ann. Math.* **43**, 351–69.

Erdélyi, A. *et al.* (1953). *Higher Transcendental Functions*, **1**. New York: McGraw-Hill.

Ihm, P. (1959). Numerical evaluation of certain multivariate normal integrals. *Sankhyā*, **21**, 363–6.

* Most of the analysis in this appendix is due to Prof. I. Marx.

JOHN, S. (1959). On the evaluation of the probability integral of a multivariate normal distribution. *Sankhyā*, **21**, 367–70.

KAMAT, A. R. (1953). Incomplete and absolute moments of the multivariate normal distribution with some applications. *Biometrika*, **40**, 21–34.

KENDALL, M. G. (1941). Proof of relations connected with the tetrachoric series and its generalization. *Biometrika*, **32**, 196–8.

KENDALL, M. G. (1945). On the analysis of oscillatory time-series. *J.R. Statist. Soc.* **108**, 93–141.

KENDALL, M. G. & STUART, A. (1958). *The Advanced Theory of Statistics*, **1**. London: Charles Griffin and Co. Ltd.

MAGNUS, W. & OBERHETTINGER, F. (1954). *Formulas and Theorems for the Functions of Mathematical Physics*. New York: Chelsea Publishing Company.

MCFADDEN, J. A. (1955). Urn models of correlation and a comparison with the multivariate normal integral. *Ann. Math. Statist.* **26**, 478–89.

MCFADDEN, J. A. (1956). An approximation for the symmetric, quadrivariate normal integral. *Biometrika*, **43**, 206–7.

MCFADDEN, J. A. (1960). An expansion for the quadrivariate normal integral for a stationary Markov process; An expansion for the quadrivariate normal integral when $\rho_{13} = \rho_{14} = \rho_{24} = 0$ (abstracts). *Ann. Math. Statist.* **31**, 526–7.

MORAN, P. A. P. (1948). Rank correlation and product-moment correlation. *Biometrika*, **35**, 203–6.

MORAN, P. A. P. (1956). The numerical evaluation of a class of integrals. *Proc. Camb. Phil. Soc.* **52**, 230–3.

PLACKETT, R. L. (1954). A reduction formula for normal multivariate integrals. *Biometrika*, **41**, 351–60.

RICE, S. O. (1944–45). Mathematical analysis of random noise. *Bell Syst. Tech. J.* **23**, 282–332; **24**, 46–156.

RICE, S. O. (1958). Distribution of the duration of fades in radio transmission: Gaussian noise model. *Bell Syst. Tech. J.* **37**, 581–635.

RUBEN, H. (1954). On the moments of order statistics in samples from normal populations. *Biometrika*, **41**, 200–27.

SCHLÄFLI, L. (1858, 1860). On the multiple integral $\int^n dx dy \dots dz$ whose limits are

$$p_1 = a_1 x + b_1 y + \dots + h_1 z > 0, \quad p_2 > 0, \quad \dots, \quad p_n > 0 \quad \text{and} \quad x^2 + y^2 + \dots + z^2 < 1.$$

Quart. J. Pure Appl. Math. **2**, 269–301; **3**, 54–68, 97–108.

SHANKS, D. (1955). Non-linear transformations of divergent and slowly convergent sequences. *J. Math. Phys.* **34**, 1–42.

VAN DER VAART, H. R. (1955). The content of some classes of non-Euclidean polyhedra for any number of dimensions, with several applications. *Kon. Nederl. Akad. Wetensch. (Amsterdam)*, A, **58**, 199–221.

Biometrika (1960), **47**, 3 *and* 4, *p.* 335
Printed in Great Britain

On a problem connected with quadratic regression*

BY R. G. LAHA AND E. LUKACS
The Catholic University of America, Washington, D.C.

1. *Introduction.* We consider two random variables X and Y and denote by $E(Y|X)$ the conditional expectation of Y given X. We say that Y has polynomial regression of order k on X if the relation

$$E(Y|X) = \beta_0 + \beta_1 X + \ldots + \beta_k X^k \tag{1·1}$$

holds almost everywhere. We assume that the first moment of Y and the moment of order k of X exist. It follows from (1·1) that

$$E(Y) = \beta_0 + \beta_1 E(X) + \ldots + \beta_k E(X^k). \tag{1·2}$$

If $k = 2$ and $\beta_2 \neq 0$ ($k = 1$ and $\beta_1 \neq 0$) then we speak of quadratic (linear) regression. If $k = 0$, that is if $E(Y|X) = E(Y)$ almost everywhere, then we say that Y has constant regression on X. The coefficients $\beta_0, \beta_1, \ldots, \beta_k$ are called the regression coefficients.

Let $X_1, X_2, \ldots, X_n$ be a sample of size n (independently and identically distributed random variables) from a population with distribution function $F(x)$. We write

$$\Lambda = X_1 + X_2 + \ldots + X_n = n\bar{X}$$

for the sum of the observations and $S = S(X_1, X_2, \ldots, X_n)$ for another statistic. In many cases we know that it is possible to find a statistic S which has constant regression on Λ. Conversely, this property determines sometimes the population.

In the present paper we consider a quadratic statistic

$$Q = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} X_i X_j + \sum_{j=1}^{n} b_j X_j$$

and study all the populations which have the property that Q has quadratic regression on Λ. It will be necessary to distinguish several cases which are defined in terms of relations between the coefficients a_{ij} and b_j of Q and the regression coefficients β_0, β_1, β_2. In each of these cases we show that the population is characterized by the property we mentioned.

We note that we consider here only a quadratic statistic which does not reduce to a linear form since it can be shown easily that every linear form $\sum_{j=1}^{n} b_j X_j$ has linear regression on Λ.

In § 2 we derive a fundamental lemma concerning polynomial regression; in § 3 we obtain a differential equation for the characteristic function of the population. Sections 4 and 5 deal with the solutions of this equation and contain the results.

2. *Two lemmas.* We give first a condition on polynomial regression.

LEMMA 1. Let X and Y be two random variables and assume that the expectations $E(Y)$ and $E(X^k)$ exist where k is a non-negative integer. The random variable Y has polynomial regression of order k on X if, and only if, the relation

$$E(Ye^{itX}) = \sum_{v=0}^{k} \beta_v E(X^v e^{itX}) \tag{2·1}$$

holds for all real t.

* This work was supported by the National Science Foundation.

If one multiplies (1·1) by e^{itX} and takes the expectation then one sees immediately that condition (2·1) is necessary.

To prove the sufficiency of the condition we assume that (2·1) is valid for all real t. Then

$$E\left\{e^{itX}\left[Y-\sum_{v=0}^{k}\beta_v X^v\right]\right\}=0,$$

or

$$\int_{-\infty}^{\infty}e^{itx}E\left(Y-\sum_{v=0}^{k}\beta_v x^v|x\right)dF_1(x)=0.$$

Here $F_1(x)$ is the marginal distribution of X. We introduce here the probability function $P_x(A)$ of the random variable X instead of the distribution function $F_1(x)$. This is a set function defined on all Borel sets of R_1. The preceding equation becomes then

$$\int_{R_1}e^{itx}E\left(Y-\sum_{v=0}^{k}\beta_v x^v|x\right)dP_x=0.$$

Let $\mu(A)=\int_A E\left(Y-\sum_{v=0}^{k}\beta_v x^v|x\right)dP_x$. This is a function of bounded variation which is defined on all Borel sets A of R_1 and we see that

$$\int_{R_1}e^{itx}d\mu=0.$$

Since the uniqueness theorem for characteristic functions is valid for the Fourier transforms of functions of bounded variation we conclude that $\mu(A)=\mu(R)=0$ for all Borel sets A. This is only possible if $E\left(Y-\sum_{v=0}^{k}\beta_v x^v|x\right)=0$ almost everywhere, so that the lemma is proven.

We next prove a lemma which is needed in § 5.

LEMMA 2. Let a, ρ and λ be three real numbers and suppose that $\rho>0$. The function

$$f(t)=[\cosh at+i\lambda\sinh at]^{-\rho}$$

is then an infinitely divisible characteristic function.

The statement of the lemma is trivial if $a=0$, it is therefore no restriction to assume that $a\neq 0$.

We consider the function

$$\gamma(z)=\cosh az+i\lambda\sinh az \tag{2·2}$$

of the complex variable $z=t+iy$ (t, y real). We see from (2·2) that $\gamma(z)$ is an entire function of order 1; the zeros of $\gamma(z)$ are the solutions of the system of equations

$$(e^{at}+e^{-at})(\cos ay-\lambda\sin ay)=0, \tag{2·3·1}$$

$$(e^{at}-e^{-at})(\sin ay+\lambda\cos ay)=0. \tag{2·3·2}$$

Equation (2·3·1) has the solutions

$$y_k=\frac{1}{a}\left(\arctan\frac{1}{\lambda}+k\pi\right)\quad(k=0,\pm1,\pm2,\ldots), \tag{2·4}$$

while the only real solution of (2·3·2) which is compatible with (2·4) is $t=0$. The zeros of $\gamma(z)$ are therefore given by

$$z_k=iy_k=\frac{i}{a}\left(\arctan\frac{1}{\lambda}+k\pi\right)\quad(k=0,\pm1,\pm2,\ldots) \tag{2·5}$$

and are thus purely imaginary. We note that $\gamma(z)$ is real for purely imaginary values of the argument. It follows then from Hadamard's factorization theorem that

$$\gamma(t) = e^{i\delta t} \prod_k \left(1 + \frac{it}{y_k}\right) \exp\left(-\frac{it}{y_k}\right), \tag{2·6}$$

where δ is a real constant and where the y_k are given by (2·4).

We conclude from P. Lévy's continuity theorem that $1/\gamma(t)$ is a characteristic function, as a limit of infinitely divisible characteristic functions $1/\gamma(t)$ is necessarily infinitely divisible so that the statement of the lemma is established.

3. *The differential equation for the characteristic function.* Let $X_1, X_2, \ldots, X_n$ be a sample of size n from a population with distribution function $F(x)$ and assume that the second moment of $F(x)$ exists. We denote by

$$f(t) = \int_{-\infty}^{\infty} e^{itx}\, dF(x)$$

the characteristic function of $F(x)$. We consider a quadratic statistic

$$Q = \sum_{i,j=1}^{n} a_{ij} X_i X_j + \sum_{j=1}^{n} b_j X_j$$

and suppose that Q has quadratic regression on $\Lambda = \sum_{j=1}^{n} X_j$ so that

$$E(Q|\Lambda) = \beta_0 + \beta_1 \Lambda + \beta_2 \Lambda^2$$

almost everywhere. It follows then from Lemma 1 that

$$E(Q e^{it\Lambda}) = \beta_0 E(e^{it\Lambda}) + \beta_1 E(\Lambda e^{it\Lambda}) + \beta_2 E(\Lambda^2 e^{it\Lambda}) \tag{3·1}$$

holds for all real t.

We denote

$$A_1 = \sum_{i=1}^{n} a_{ii}, \quad A_2 = \sum_{\substack{i,j=1 \\ i \neq j}}^{n} a_{ij}, \quad B = \sum_{j=1}^{n} b_j.$$

In a certain neighbourhood of the origin $f(t)$ is different from zero so that we can introduce $\phi(t) = \ln f(t)$. Then

$$\frac{f'(t)}{f(t)} = \phi'(t) \quad \text{and} \quad \frac{f''(t)}{f(t)} = \phi''(t) + [\phi'(t)]^2.$$

We obtain, by means of some elementary computations, the relations

$$E(Q e^{it\Lambda}) = -[f(t)]^n \{A_1 \phi''(t) + (A_1 + A_2)[\phi'(t)]^2 + Bi\phi'(t)\} \tag{3·2·1}$$

and

$$\beta_0 E(e^{it\Lambda}) + \beta_1 E(\Lambda e^{it\Lambda}) + \beta_2 E(\Lambda^2 e^{it\Lambda}) = -[f(t)]^n \{n\beta_2 \phi''(t) + n^2 \beta_2 [\phi'(t)]^2 + n\beta_1 i\phi'(t) - \beta_0\}. \tag{3·2·2}$$

These relations are valid in the neighbourhood of the origin in which $\phi(t)$ is defined. We see from equations (3·2·1), (3·2·2) and (3·1) that

$$\gamma_1 \phi''(t) + \gamma_2 [\phi'(t)]^2 + i\gamma_3 \phi'(t) = \beta_0, \tag{3·3}$$

where

$$\gamma_1 = n\beta_2 - A_1, \quad \gamma_2 = n^2\beta_2 - A_1 - A_2, \quad \gamma_3 = n\beta_1 - B. \tag{3·4}$$

It is convenient to introduce the function

$$\theta(t) = (1/i)\, \phi'(t). \tag{3·5}$$

Equation (3·3) can then be written as

$$i\gamma_1 \frac{d\theta}{dt} = \gamma_2\theta^2 + \gamma_3\theta + \beta_0, \tag{3·6}$$

where we wrote θ for $\theta(t)$. It follows from (3·5) that

$$\theta(0) = (1/i)\,\phi'(0) = \alpha \tag{3·6·1}$$

and

$$\left.\frac{d\theta}{dt}\right|_{t=0} = \frac{1}{i}\phi''(0) = i\sigma^2. \tag{3·6·2}$$

Here α and σ^2 are the mean and the variance respectively of the distribution $F(x)$. We put $t = 0$ in (3·6) and obtain the relation

$$-\sigma^2\gamma_1 = \gamma_2\alpha^2 + \gamma_3\alpha + \beta_0 \tag{3·7}$$

between the coefficients of Q and the regression coefficients. It would also have been possible to obtain (3·7) from equation (1·2). This is a consistency condition which must be satisfied.

Our next aim is to obtain all the solutions of equation (3·6). For this investigation we must consider several cases.

We need not concern ourselves with the possibility that all coefficients γ_1, γ_2, γ_3 vanish. If $\gamma_1 = 0$ while at least one of the coefficients γ_2 and γ_3 is different from zero then we see that $\theta(t)$ and $\phi'(t)$ are constants so that we obtain a degenerate distribution. We can, therefore, assume without loss of generality that $\gamma_1 \neq 0$.

In § 4 we discuss the case $\gamma_2 = 0$. Here we have to study separately the cases where

$$\gamma_1 \neq 0, \quad \gamma_2 = 0, \quad \gamma_3 = 0 \tag{3·8}$$

and

$$\gamma_1 \neq 0, \quad \gamma_2 = 0, \quad \gamma_3 \neq 0. \tag{3·9}$$

In § 5 we deal with the case $\gamma_2 \neq 0$ and must distinguish three possibilities which depend on the sign of the discriminant $\Delta = \gamma_3^2 - 4\gamma_2\beta_0$.

4. *The case* $\gamma_2 = 0$. We first investigate the case (3·8) and prove the following theorem.

THEOREM 1. Let $X_1, X_2, \ldots, X_n$ be a sample of size n taken from a population which has a finite variance σ^2. Consider a quadratic statistic

$$Q = \sum_{i,j=1}^{n} a_{ij}X_iX_j + \sum_{j=1}^{n} b_jX_j$$

such that

$$A = A_1(n-1) - A_2 \neq 0, \tag{4·1}$$

where

$$A_1 = \sum_{i=1}^{n} a_{ii}, \quad A_2 = \sum_{\substack{i,j=1 \\ i\neq j}}^{n} a_{ij}.$$

Let β_1 and β_2 be two real constants such that

$$\beta_1 = \frac{B}{n} \quad \left(\text{where } B = \sum_{j=1}^{n} b_j\right), \tag{4·2·1}$$

$$\beta_2 = \frac{1}{n^2}(A_1 + A_2). \tag{4·2·2}$$

The relation

$$E(Q|\Lambda) = \beta_0 + \beta_1\Lambda + \beta_2\Lambda^2 \tag{4·3}$$

holds almost everywhere if, and only if, the following two conditions are satisfied

$$\text{(i)}\quad \beta_0 = A\frac{\sigma^2}{n}; \qquad \text{(ii)}\quad \text{the population is normal.}$$

We first note that the relations (4·1), (4·2·1) and (4·2·2) are equivalent to (3·8). We first prove that the conditions (i) and (ii) are necessary and assume therefore that (4·3) is satisfied. The differential equation (3·6) for $\theta(t)$ reduces in view of (3·8) to

$$i\gamma_1\frac{d\theta}{dt} = \beta_0, \tag{4·4}$$

we put here $t = 0$ and use (3·6·2) to show that $\beta_0 = -\gamma_1\sigma^2$. Condition (i) follows then immediately from (3·4) and (4·2·2). We integrate equation (4·4) for $\theta(t)$ with the initial conditions (3·6·1) and (3·6·2) and obtain

$$\phi(t) = -\tfrac{1}{2}\sigma^2t^2 + i\alpha t,$$

or

$$f(t) = \exp[-\tfrac{1}{2}\sigma^2t^2 + i\alpha t].$$

This is the characteristic function of a normal (possibly degenerate) distribution so that (ii) follows.

To prove the sufficiency of condition (4·3) we assume that the population is normal so that

$$\phi(t) = \log f(t) = -\tfrac{1}{2}\sigma^2t^2 + i\alpha t.$$

It follows then from (3·2·1) and (3·2·2) that

$$E(Q\,e^{it\Lambda}) = -[f(t)]^n\{\sigma^4(A_1+A_2)t^2 - \sigma^2[2\alpha(A_1+A_2)+B]it - A_1\sigma^2 - B\alpha - \alpha^2(A_1+A_2)\} \tag{4·5·1}$$

and

$$\beta_0E(e^{it\Lambda}) + \beta_1E(\Lambda\,e^{it\Lambda}) + \beta_2E(\Lambda^2\,e^{it\Lambda})$$
$$= -[f(t)]^n\{\sigma^4n^2\beta_2t^2 - \sigma^2[2\alpha n^2\beta_2 + \beta_1 n]it - \beta_0 - \alpha\beta_1 n - n\beta_2\sigma^2 - n^2\alpha^2\beta_2\}. \tag{4·5·2}$$

In view of (4·2·1), (4·2·2) and (i) we see from (4·5·1) and (4·5·2) that

$$E(Q\,e^{it\Lambda}) = \beta_0E(e^{it\Lambda}) + \beta_1E(\Lambda\,e^{it\Lambda}) + \beta_2E(\Lambda^2\,e^{it\Lambda})$$

for all real t. We conclude then from Lemma 1 that (4·3) holds almost everywhere. We must examine the case where (3·9) holds and prove the following theorem.

THEOREM 2. Let $X_1, X_2, \ldots, X_n$ be a sample of size n taken from a population which has finite variance σ^2. Consider a quadratic statistic

$$Q = \sum_{i,j=1}^{n} a_{ij}X_iX_j + \sum_{j=1}^{n} b_jX_j$$

such that condition (4·1) of Theorem 1 holds. Let β_1 and β_2 be two real constants such that

$$\beta_1 \neq \frac{B}{n} \quad \left(\text{where } B = \sum_{j=1}^{n} b_j\right), \tag{4·6·1}$$

$$\beta_2 = \frac{1}{n^2}(A_1 + A_2). \tag{4·6·2}$$

The relation

$$E(Q|\Lambda) = \beta_0 + \beta_1\Lambda + \beta_2\Lambda^2 \tag{4·3}$$

holds almost everywhere if, and only if, the following three conditions are satisfied:

(i) the population has the Poisson type characteristic function

$$f(t) = \exp[\lambda(e^{i\rho t} - 1) + i\mu t],$$

where $\lambda > 0$, $\rho \neq 0$ and μ are three real constants,

$$\text{(ii)} \quad \beta_0 = -\frac{A\rho\mu}{n}, \qquad \text{(iii)} \quad \beta_1 = \frac{1}{n^2}(A\rho + nB).$$

Here the relations (4·1), (4·6·1) and (4·6·2) are equivalent to (3·9). To prove the necessity of conditions (i), (ii) and (iii) we assume that (4·3) is satisfied. The differential equation (3·6) for $\theta(t)$ becomes then

$$i\gamma_1 \frac{d\theta}{dt} = \gamma_3\theta + \beta_0. \tag{4·7}$$

We integrate equation (4·7) with the initial conditions (3·6·1) and (3·6·2) and obtain

$$\phi(t) = \frac{\sigma^2\gamma_1^2}{\gamma_3^2}\{\exp(-i\gamma_3 t/\gamma_1) - 1\} - \frac{\beta_0}{\gamma_3} it.$$

We write

$$\lambda = \frac{\sigma^2\gamma_1^2}{\gamma_3^2}, \tag{4·8·1}$$

$$\rho = -\frac{\gamma_3}{\gamma_1}, \tag{4·8·2}$$

$$\mu = -\frac{\beta_0}{\gamma_3} \tag{4·8·3}$$

and obtain the Poisson-type characteristic function (i). We see then easily from (3·4), (4.6.2), (4.8.2) and (4.8.3) that (ii) and (iii) are satisfied.

To prove the sufficiency of (4·3) we assume that the characteristic function of the population is given by (i) and that β_0 and β_1 are defined by (ii) and (iii). We see from (3·2·1) and (3·2·2) that

$$E(Q\,e^{it\Lambda}) = [f(t)]^n\{\rho^2\lambda^2(A_1+A_2)\,e^{2i\rho t} + \rho\lambda[\rho A_1 + 2\mu(A_1+A_2) + B]\,e^{i\rho t} + \mu^2(A_1+A_2) + B\mu\}, \tag{4·9·1}$$

$$\beta_0 E(e^{it\Lambda}) + \beta_1 E(\Lambda\, e^{it\Lambda}) + \beta_2 E(\Lambda^2 e^{it\Lambda}) = [f(t)]^n\{\rho^2\lambda^2 n^2\beta_2\, e^{2i\rho t} + \rho\lambda[n\beta_2 + 2\mu n^2\beta_2 + \beta_1 n]\, e^{i\rho t} + \beta_0 + \beta_1 n\mu + \mu^2 n^2\beta_2\}. \tag{4·9·2}$$

We use the relations (4·6·2), (ii) and (iii) and Lemma 1 to show that (4·3) holds almost everywhere.

5. *The case* $\gamma_2 \neq 0$. Our next theorem deals with the case where $\gamma_2 \neq 0$ while

$$\Delta = \gamma_3^2 - 4\gamma_2\beta_0 > 0.$$

THEOREM 3. Let $X_1, X_2, \ldots, X_n$ be a sample of size n taken from a population with finite variance σ^2. Consider a quadratic statistic

$$Q = \sum_{i,j=1}^{n} a_{ij} X_i X_j + \sum_{j=1}^{n} b_j X_j$$

such that condition (4·1) holds. Let β_0, β_1 and β_2 be three real constants such that

$$\beta_2 \neq \frac{A_1 + A_2}{n^2}, \tag{5·1·1}$$

$$(n\beta_1 - B)^2 > 4\beta_0(n^2\beta_2 - A_1 - A_2), \tag{5·1·2}$$

where

$$B = \sum_{j=1}^{n} b_j.$$

The relation

$$E(Q|\Lambda) = \beta_0 + \beta_1 \Lambda + \beta_2 \Lambda^2 \tag{4·3}$$

holds almost everywhere if, and only if, the following four conditions are satisfied:

(i) The population has the characteristic function

$$f(t) = [p\, e^{it\mu_1} + q\, e^{it\mu_2}]^\rho \quad (p+q=1),$$

(ii) $\beta_0 = -\dfrac{\rho^2 A \mu_1 \mu_2}{n\rho - 1}$, (iii) $\beta_1 = \dfrac{B}{n} + \dfrac{\rho A(\mu_1+\mu_2)}{n(n\rho-1)}$, (iv) $\beta_2 = \dfrac{A_1(\rho-1)+A_2\rho}{n(n\rho-1)}$.

Thus, the population is either a binomial population (if ρ is a positive integer, then necessarily $p > 0$, $q > 0$) or a negative binomial population (if $\rho < 0$ then $pq < 0$).

We prove first that condition (4·3) is necessary. It follows from (3·4), (5·1·1) and (5·1·2) that $\gamma_2 \neq 0$ and $\Delta = \gamma_3^2 - 4\gamma_2\beta_0 > 0$. Since Δ is the discriminant of the quadratic form $\gamma_2\theta^2 + \gamma_3\theta + \beta_0$ one can write the differential equation (3·6) for $\theta(t)$ in the form

$$i\gamma_1 \frac{d\theta}{dt} = \gamma_2(\theta - \eta_1)(\theta - \eta_2), \tag{5·2}$$

where η_1 and η_2 are both real and $\eta_1 \neq \eta_2$. We integrate first this equation, then the equation which results from (3·5) and use the initial conditions (3·6·1) and (3·6·2). In this way we obtain

$$\phi(t) = \frac{\gamma_1}{\gamma_2} \ln \left\{ \frac{\alpha - \eta_2}{\eta_1 - \eta_2} \exp\left(\frac{\eta_1\gamma_2}{\gamma_1} it\right) - \frac{\alpha - \eta_1}{\eta_1 - \eta_2} \exp\left(\frac{\eta_2\gamma_2}{\gamma_1} it\right) \right\}.$$

If we write $f(t) = \exp[\phi(t)]$ and $p = \dfrac{\alpha - \eta_2}{\eta_1 - \eta_2}$, $q = -\dfrac{\alpha - \eta_1}{\eta_1 - \eta_2}$,

$$\mu_1 = \frac{\eta_1\gamma_2}{\gamma_1}, \quad \mu_2 = \frac{\eta_2\gamma_2}{\gamma_1}, \quad \rho = \frac{\gamma_1}{\gamma_2}$$

we see that $f(t)$ has the form specified in (i).

We put $t = 0$ in (5·2) and see, using (3·6·1) and (3·6·2) that $-\sigma^2\rho = (\alpha - \eta_1)(\alpha - \eta_2)$. From this it follows easily that $p > 0$, $q > 0$ in case $\rho > 0$ while $pq < 0$ if $\rho < 0$. We note that

$$\eta_1 + \eta_2 = -\frac{\gamma_3}{\gamma_2} \quad \text{and} \quad \eta_1\eta_2 = \frac{\beta_0}{\gamma_2}.$$

Using these relations and the formulae defining μ_1, μ_2 and ρ as well as (3·4) we can derive the expressions (ii), (iii) and (iv) for β_0, β_1 and β_2.

To prove that our conditions are sufficient we assume that the population has the characteristic function specified by (i) and that β_0, β_1 and β_2 are given by (ii), (iii) and (iv), respectively. We proceed then as in the proof of Theorem 2, and can show that (4·3) holds almost everywhere.

We consider next the case where $\Delta = \gamma_3^2 - 4\gamma_2\beta_0 = 0$ and prove the following theorem.

THEOREM 4. Let $X_1, X_2, \ldots, X_n$ be a sample of size n taken from a population with finite variance σ^2. Consider a quadratic statistic

$$Q = \sum_{i,j=1}^{n} a_{ij} X_i X_j + \sum_{j=1}^{n} b_j X_j$$

such that condition (4·1) holds. Let β_0, β_1 and β_2 be three real constants such that

$$\beta_2 \neq \frac{A_1+A_2}{n^2}, \tag{5·3·1}$$

$$(n\beta_1-B)^2 = 4\beta_0(n^2\beta_2-A_1-A_2), \tag{5.3.2}$$

where $$B = \sum_{j=1}^{n} b_j.$$

The relation $$E(Q|\Lambda) = \beta_0+\beta_1\Lambda+\beta_2\Lambda^2 \tag{4·3}$$

holds almost everywhere if, and only if, the following four conditions are satisfied:

(i) The population is a Gamma population with characteristic function

$$f(t) = e^{i\mu}\left(1-\frac{it}{a}\right)^{-\rho} \quad (\rho > 0,\ a \neq 0),$$

(ii) $\beta_0 = \dfrac{\mu^2 A}{n\rho+1}$, (iii) $\beta_1 = \dfrac{B}{n} - \dfrac{2\mu A}{n(n\rho+1)}$, (iv) $\beta_2 = \dfrac{A_1(\rho+1)+A_2\rho}{n(n\rho+1)}$.

We see from (3·4), (5·3·1) and (5·3·2) that $\gamma_2 \neq 0$ and that $\Delta = \gamma_3^2 - 4\gamma_2\beta_0 = 0$. Since Δ is the discriminant of the quadratic from $\gamma_2\theta^2+\gamma_3\theta+\beta_0$ we can write the differential equation (3·6) for $\theta(t)$ in the form

$$i\gamma_1\frac{d\theta}{dt} = \gamma_2(\theta-\eta)^2, \tag{5·4}$$

where η is real. To prove that condition (4·3) is necessary we solve this equation and determine then $f(t)$ and obtain

$$f(t) = e^{i\eta t}\left[1+\frac{(\alpha-\eta)\gamma_2}{\gamma_1}it\right]^{\gamma_1/\gamma_2}.$$

We put $t = 0$ in (5·4) and see that

$$\frac{\gamma_1}{\gamma^2} = -\frac{(\alpha-\eta)^2}{\sigma^2} < 0.$$

If we write $$\rho = -\frac{\gamma_1}{\gamma_2}, \quad a = -\frac{\gamma_1}{\gamma_2(\alpha-\eta)}, \quad \mu = \eta,$$

then we see that $f(t)$ has the form specified in (i). We note that

$$\frac{\beta_0}{\gamma_2} = \eta^2 \quad \text{and} \quad \frac{\gamma_3}{\gamma_2} = -2\eta.$$

From these relations and from (3·4) we get easily the expressions (ii), (iii) and (iv) for β_0, β_1 and β_2.

The proof of the sufficiency of our conditions is given in the same way as in the preceding theorems and is therefore omitted.

Finally, we consider the case where $\Delta = \gamma_3^2 - 4\gamma_2\beta_0 < 0$ and prove the following theorem.

THEOREM 5. Let $X_1, X_2, \ldots, X_n$ be a sample of size n taken from a population with variance σ^2. Consider a quadratic statistic

$$Q = \sum_{i,j=1}^{n} a_{ij}X_iX_j + \sum_{j=1}^{n} b_jX_j$$

such that condition (4·1) holds. Let β_0, β_1 and β_2 be three constants such that

$$\beta_2 \neq \frac{A_1+A_2}{n^2}, \tag{5·5·1}$$

$$(n\beta_1-B)^2 < 4\beta_0(n^2\beta_2-A_1-A_2), \tag{5·5·2}$$

where $$B = \sum_{j=1}^{n} b_j.$$

The relation $$E(Q|\Lambda) = \beta_0+\beta_1\Lambda+\beta_2\Lambda^2 \tag{4·3}$$

holds almost everywhere if, and only if, the following four conditions are satisfied:

(i) the characteristic function of the population is given by

$$f(t) = e^{i\mu t}[\cosh at + i\lambda \sinh at]^{-\rho},$$

where a, λ, μ, ρ are real constants and $\rho > 0$, $a \neq 0$,

(ii) $\beta_0 = \dfrac{A(\mu^2+\rho^2a^2)}{n\rho+1}$, (iii) $\beta_1 = \dfrac{B}{n} - \dfrac{2A\mu}{n(n\rho+1)}$, (iv) $\beta_2 = \dfrac{A_1(1+\rho)+\rho A_2}{n(n\rho+1)}$.

We see from (3·4), (5·5·1) and (5·5·2) that $\gamma_2 \neq 0$ and $\Delta < 0$. We can then write the differential equation (3·6) for $\theta(t)$ in the form

$$i\gamma_1\frac{d\theta}{dt} = \gamma_2(\theta-\eta)(\theta-\overline{\eta}), \tag{5·6}$$

where $\eta = \mu+i\nu$, $\overline{\eta} = \mu-i\nu$ (μ, ν real). We determine $f(t)$ from this equation and obtain

$$f(t) = e^{i\mu t}\left[\cosh\frac{\nu\gamma_2}{\gamma_1}t + i\lambda\sinh\frac{\nu\gamma_2}{\gamma_1}t\right]^{\gamma_1/\gamma_2}.$$

We put $t = 0$ in (5·6) and see that

$$\frac{\gamma_1}{\gamma_2} = -\frac{(\alpha-\mu)^2+\nu^2}{\sigma^2} < 0.$$

We write $$\rho = -\frac{\gamma_1}{\gamma_2} > 0, \quad a = \frac{\nu\gamma_2}{\gamma_1}$$

and see that $f(t)$ has the form specified in (i). We note that

$$\eta+\overline{\eta} = -\frac{\gamma_3}{\gamma_2} \quad \text{and} \quad \eta\overline{\eta} = \frac{\beta_0}{\gamma_2}.$$

From these relations and from (3·4) we get easily (ii), (iii), and (iv).

The proof of the sufficiency of our conditions is given in the same way as in Theorems 1, 2 and 3.

In conclusion we note that Tweedie (1946) considered earlier the regression of the sample variance on the sample mean. He obtained particular cases of some of the theorems of the present paper. The authors are indebted to Mr Tweedie for calling their attention to his paper.

Reference

Tweedie, M. C. K. (1946). The regression of the sample variance on the sample mean. *J. Lond. Math. Soc.* **21**, 22–8.

Biometrika (1960), **47**, 3 *and* 4, *p.* 345
With 1 *text-figure*
Printed in Great Britain

Small sample behaviour of certain tests of the hypothesis of equal means under variance heterogeneity*

BY ROGER S. McCULLOUGH, JOHN GURLAND
Department of Statistics, Iowa State University
AND LLOYD ROSENBERG
Department of Mathematical Statistics, Columbia University

SUMMARY

By using a finite series representation of the distribution of a certain class of statistics, and by utilizing other techniques presented here, two types of statistic are developed for testing the hypothesis of the equality of the means of two normal populations using small sample sizes. One type of statistic, called unilateral, effectively controls the size of the test under variance heterogeneity if it is known *a priori* that the variance of one specified population is greater than that of the other. The second type of statistic, called bilateral, controls the size when there is no *a priori* knowledge of the population variances. Tables are included specifying both types of statistic and giving size and power.

1. INTRODUCTION

Suppose X_1 and X_2 are two independent Normal random variables with means μ_1 and μ_2 and variances σ_1^2 and σ_2^2, respectively. Samples of sizes n_1 and n_2 drawn from the corresponding populations are denoted by $x_{ij}\,(i = 1, 2; j = 1, 2, \ldots, n_i)$. It is desired to test the hypothesis H_0: $\mu_1 = \mu_2$ (without assuming $\sigma_1^2 = \sigma_2^2$), and the problem is to find a statistic for making this test which has optimal properties in some sense. The usual functions of the samples are defined as follows

$$\bar{x}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} x_{ij}, \quad \Sigma_i = \sum_{j=1}^{n_i}(x_{ij}-\bar{x}_i)^2. \qquad (i = 1, 2). \tag{1}$$

Also let the parameter R denote the ratio σ_1^2/σ_2^2 of the true population variances. R can, of course, take on values from zero to infinity.

The effect of variance heterogeneity on the distributions of certain statistics used in testing H_0 will be investigated for the case of small sample sizes 3, 5, 7 and 9. These sample sizes are convenient in the application of the finite series expansion employed here, but larger sample sizes could also be treated by the same technique.

Welch (1937, 1947) has considered the above problem, and the following statistics for use in testing H_0 are discussed in his second paper. They are squared here for simplicity.

$$u^2 = \frac{(\bar{x}_1-\bar{x}_2)^2}{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\frac{\Sigma_1+\Sigma_2}{n_1+n_2-2}}, \tag{2}$$

$$v^2 = \frac{(\bar{x}_1-\bar{x}_2)^2}{\frac{\Sigma_1}{n_1(n_1-1)}+\frac{\Sigma_2}{n_2(n_2-1)}}, \tag{3}$$

* This research was monitored by the Air Force Office of Scientific Research and Development Command.

$$z^2 = \frac{(\bar{x}_1 - \bar{x}_2)^2}{\frac{\Sigma_1}{n_1(n_1-3)} + \frac{\Sigma_2}{n_2(n_2-3)}} \quad (n_1 > 3, n_2 > 3). \tag{4}$$

Chand (1950), Gronow (1951) and Hsu (1938) have examined some of the properties of the u and v statistics and have tabulated size and power for a few special cases.

It was suggested in Welch (1937) that a statistic of the general form

$$Y(r_1, r_2) = \frac{(\bar{x}_1 - \bar{x}_2)^2}{r_1\Sigma_1 + r_2\Sigma_2}, \tag{5}$$

where r_1 and r_2 are constants appropriately chosen, might be considered in attempting to control the size of the test for different values of $R = \sigma_1^2/\sigma_2^2$. It is obviously possible so to choose r_1 and r_2 that the hypothesis H_0 is rejected only when $Y(r_1, r_2) > 1$; this choice will always be made in the present paper. The distribution of (5) under H_0 is that of

$$\frac{\chi_1^2}{\lambda_1\chi_{f_1}^2 + \lambda_2\chi_{f_2}^2}, \tag{6}$$

where $$f_i = n_i - 1, \quad \lambda_1 = \frac{Rr_1}{\frac{R}{n_1} + \frac{1}{n_2}} \quad \text{and} \quad \lambda_2 = \frac{r_2}{\frac{R}{n_1} + \frac{1}{n_2}},$$

and all the χ^2 variables are independent. Under the alternative hypothesis $\mu_1 \neq \mu_2$, the distribution of (5) is that of

$$\frac{\chi'^2_{1,\lambda}}{\lambda_1\chi_{f_1}^2 + \lambda_2\chi_{f_2}^2}, \tag{7}$$

where the variables are independent, and the numerator is a non-central χ^2 with one degree of freedom, and non-centrality λ given by

$$\lambda = \frac{(\mu_1 - \mu_2)^2}{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \left(\frac{\delta}{\sigma_2}\right)^2 \frac{n_1 n_2}{n_1 + Rn_2}, \tag{8}$$

where $\delta = \mu_1 - \mu_2$.

Our attack on the problem is similar to that of Welch (1937) but differs from that of Welch (1947). We consider the class of statistics (5) and examine the size and power for a fixed critical point (namely unity). In Welch (1947) the corresponding critical points are not fixed but are functions of the sample variances. Ura (1955) has studied the power of the test considered by Welch (1947) and elaborated by Aspin (1949), in the case of equal sample sizes. Wald (1955) has proposed a statistic in the case of equal sample sizes which also uses a random critical point that is a function of the sample variances.

In the present article we treat separately two kinds of statistic, those with unilateral and those with bilateral control of size. The first kind, called unilateral, is used if it is known *a priori* that the variance of one specified population is greater than that of the other. This statistic keeps the size of the test less than or equal to a given value over the relevant range of $R \geqslant 1$. The second kind, called bilateral, is used when there is no *a priori* knowledge of the population variances. It keeps the size less than or equal to a preassigned value over the whole range of R. For the unilateral case, all pairs of sample sizes 3, 5, 7 and 9 are considered (see Table 1), and size and power are given (see Tables 2 and 3). For the bilateral case it is shown how tables of the percentage points of Student's t may be used to find the appropriate statistics for any pair of sample sizes.

2. Techniques for evaluating the required probabilities

2·1. *Calculation of the size of the critical regions*

To calculate the size of the test using (5), it is necessary to evaluate probabilities of the form

$$P\{Y(r_1, r_2) > 1 \mid R\}, \tag{9}$$

where R is a constant. For specific values of n_1, n_2, r_1 and r_2 and for Y distributed as (6), the size given by (9) can be plotted against R. Some graphs for the case $n_1 = 3$, $n_2 = 5$ and for selected r_1 and r_2 are given in Fig. 1. It should be noted that the curves approach horizontal asymptotes as $R \to 0$ (i.e. $\sigma_1 \to 0$) and $R \to \infty$ (i.e. $\sigma_2 \to 0$). In these two extreme cases the statistic $Y(r_1, r_2)$ has the following structures.

$$\frac{\chi_1^2}{n_2 r_2 \chi_{f_2}^2} \quad \text{for} \quad R = 0 \tag{10}$$

and

$$\frac{\chi_1^2}{n_1 r_1 \chi_{f_1}^2} \quad \text{for} \quad R = \infty. \tag{11}$$

To evaluate probabilities corresponding to (10) and (11) it is convenient to transform to the Incomplete Beta Function as follows.

Let

$$G_{n,m}(c) = P\left\{\frac{\chi_n^2}{\chi_m^2} > c\right\}, \tag{12}$$

where the χ^2 variables are independent. Then

$$G_{n,m}(c) = I_x(\tfrac{1}{2}m, \tfrac{1}{2}n), \tag{13}$$

where

$$x = \frac{1}{1+c}$$

and

$$I_x(a, b) = \frac{1}{B(a, b)} \int_0^x t^{a-1}(1-t)^{b-1}\, dt.$$

The Incomplete Beta Function $I_x(a, b)$ was tabulated in Pearson (1934). If $G_{1,m}(c)$ has the value α, then c can be written as

$$c = (t_m^{\alpha})^2/m, \tag{14}$$

where t_m^{α} is the two-sided α % point of Student's t with m degrees of freedom.

In computing probabilities of the general form (9), use was made of the following theorem.

Theorem 1. Let Y be a random variable with the structure

$$Y = \frac{\chi_n^2}{\lambda_1 \chi_{f_1}^2 + \lambda_2 \chi_{f_2}^2}, \tag{15}$$

where the χ^2 variables are independent, the $f_j = 2g_j$ are even integers, and the λ_j are positive constants. Then the distribution function of Y is given by

$$F_Y(y) = \sum_{j=1}^{2} \sum_{s=1}^{g_j} \alpha_{js} F_{n,2s}(\lambda_j y), \tag{16}$$

where $F_{n,m}(x) = 1 - G_{n,m}(x)$. The constants α_{js} are given by

$$\alpha_{1s} = (-1)^{g_1 - s} \frac{\Gamma(g_1 + g_2 - s)}{\Gamma(g_2)\,\Gamma(g_1 - s + 1)} \frac{\lambda_1^{g_2} \lambda_2^{g_1 - s}}{(\lambda_1 - \lambda_2)^{g_1 + g_2 - s}} \quad (s = 1, 2, 3, \ldots, g_1). \tag{17}$$

$$\alpha_{2s} = (-1)^{g_2-s} \frac{\Gamma(g_1+g_2-s)}{\Gamma(g_1)\,\Gamma(g_2-s+1)} \frac{\lambda_1^{g_2-s}\lambda_2^{g_1}}{(\lambda_2-\lambda_1)^{g_1+g_2-s}} \quad (s = 1, 2, 3, \ldots, g_2). \tag{18}$$

The above theorem follows easily from the distribution of a linear combination of χ^2 random variables with even degrees of freedom (cf. Box (1954), Theorem 2·4). This theorem although not stated as such is available in Hsu (1938) and Chand (1950), and has been used in computing the sizes of the various tests described in §3.

2·2. *Calculation of the power of the test*

In computing probabilities required for the power of the tests proposed, use was made of the following theorem.

THEOREM 2. Let Y be a random variable of the distributional form

$$Y = \frac{\chi'^2_{n,\lambda}}{\lambda_1\chi^2_{f_1} + \lambda_2\chi^2_{f_2}}, \tag{19}$$

where the χ^2 variables are independent, and the numerator is a non-central χ^2 with non-centrality λ. Then Y has the distribution function

$$F_Y(y) = e^{-\frac{1}{2}\lambda} \sum_{j=1}^{2} \sum_{s=1}^{g_j} \alpha_{js} \sum_{r=0}^{\infty} \frac{\lambda^r}{r!\,2^r} F_{n+2r,2s}(\lambda_j y), \tag{20}$$

where notation is the same as in Theorem 1.

This theorem follows easily from Theorem 1, and has been used by Chand (1950) and indirectly by Hsu (1938).

3. DETERMINATION OF OPTIMAL STATISTICS

3·1. *Statistics with unilateral control of size*

In using the u^2, v^2 and z^2 statistics defined by (2), (3) and (4), the critical values of the tests were so chosen that the size of each test was exactly equal to the preassigned value α when $R = 1$ (i.e. $\sigma_1 = \sigma_2$).

When the case $n_1 = 3, n_2 = 5$ is used as an example it is found that Welch's condition that Ez^2 have equal values at $R = 0$ and $R = \infty$ is inapplicable. Therefore a modified z^2 statistic with $r_1 = 1{\cdot}5840$ and $r_2 = 0{\cdot}2375$, which gives $E|z|$ equal values at $R = 0$ and $R = \infty$, is used. As can be seen from the graphs in Fig. 1, the curves become flatter in the region $R > 1$ as r_1 increases and so r_2 decreases, up to a certain point. For the particular values $r_1 = 3{\cdot}0855$, $r_2 = 0{\cdot}0822$, the size has an asymptotic value of 0·05 as $R \to \infty$, and is only slightly less than 0·05 for the entire range $1 < R < \infty$. This statistic can be extremely useful if it is known that $\sigma_1^2 > \sigma_2^2$. For example, the populations might consist of measurements made by two different techniques, a particular one of which is known to be more precise than the other. Such information is utilized by this statistic in keeping the size practically constant over the relevant range of $R > 1$. Such a statistic is called a *unilateral* statistic. It can be looked upon as optimal within the class (5) because no other statistic in this class keeps the size as nearly constant and also $\leqslant \alpha$ for $R \geqslant 1$.

Unilateral statistics can be found for all combinations of sample sizes from the following two conditions

$$\left.\begin{aligned} P\{Y(r_1, r_2) > 1 \,|\, R = 1\} &= \alpha, \\ P\{Y(r_1, r_2) > 1 \,|\, R = \infty\} &= \alpha. \end{aligned}\right\} \tag{21}$$

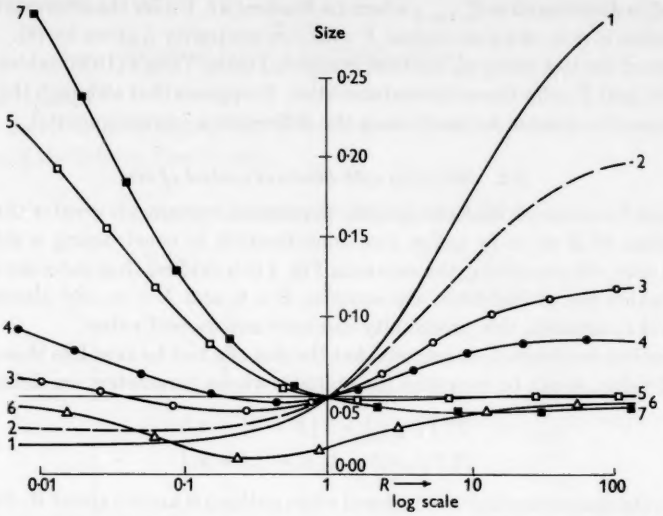


Fig. 1. Size curves corresponding to tests using several statistics (as shown in the table) for the case $n_1 = 3$, $n_2 = 5$ and $\alpha = 0{\cdot}05$.

Guide to curves

Curve number	Statistic used	r_1	r_2
1	—	0·2773	0·6932
2	u^2	·5322	·5322
3	v^2	1·1007	·3322
4	z^2 (modified)	1·5840	·2375
5	Unilateral	3·0855	·0822
6	Bilateral	3·0855	·3854
7	—	3·8690	·0387

The second equation can be written

$$P\left\{\frac{\chi_1^2}{\chi_{f_1}^2} > r_1 n_1\right\} = \alpha$$

and from this we get directly

$$r_1 = \frac{(t_{f_1}^{\alpha})^2}{n_1 f_1}. \tag{22}$$

The parameter r_2 can be written as

$$r_2 = \frac{\phi}{n_2 f_2} \tag{23}$$

and ϕ is given in Table 1 for the cases $\alpha = 0{\cdot}05$ and $\alpha = 0{\cdot}01$, and sample sizes 3, 5, 7 and 9.

In Table 3 the power is given, as in Gronow (1951), as a function of δ/σ_2 for convenience. The three values of δ/σ_2 were chosen to give power near 0·3, 0·6 and 0·9 in order to have a more extensive picture of the relationship of δ/σ_2 to power.

The power of the test using the unilateral statistic can be compared with the power of an analagous t-test in which the ratio of variances is assumed to be known. A t^2-statistic that could then be used for testing $H_0: \mu_1 = \mu_2$ is

$$u_R^2 = \frac{(\bar{x}_1 - \bar{x}_2)^2}{\dfrac{\Sigma_1/R + \Sigma_2}{n_1 + n_2 - 2}\left(\dfrac{R}{n_1} + \dfrac{1}{n_2}\right)}. \tag{24}$$

Under H_0, u_R^2 is distributed as $t^2_{n_1+n_2-2}$ where t is Student's t. Under the alternative hypothesis the distribution is that of a non-central F with non-centrality λ given by (8).

The power of the test using u_R^2 has been computed using Tang's (1938) tables for the same values of δ/σ_2 and R as for the unilateral statistics. It appears that although the power of the t-test is always the greater, in many cases the difference is inconsequential.

3·2. *Statistics with bilateral control of size*

If it cannot be assumed that one specific population variance is greater than the other, then all values of R must be taken into consideration in constructing a statistic which controls the size. On examining the curves in Fig. 1 it is evident that there are curves in the family for which the probabilities are equal at $R = 0$, and $R = \infty$. By altering the parameters r_1 and r_2 suitably, this probability can take any desired value.

Since practical considerations suggest that the size of a test be kept less than or equal to a preassigned value, it can be seen that the statistic whose parameters are defined by

$$\left.\begin{aligned} P\{Y(r_1, r_2) > 1 \,|\, R = 0\} &= \alpha, \\ P\{Y(r_1, r_2) > 1 \,|\, R = \infty\} &= \alpha. \end{aligned}\right\} \tag{25}$$

is the best in the class of statistics considered when nothing is known about R. Such statistics will be called *bilateral*. From (25) we can get values for r_1 and r_2 directly for all n_1 and n_2 and for any value of α using tables of Student's t:

$$\left.\begin{aligned} r_1 &= (t^{\alpha}_{f_1})^2/n_1 f_1 \\ r_2 &= (t^{\alpha}_{f_2})^2/n_2 f_2. \end{aligned}\right\} \tag{26}$$

4. USE OF THE TABLES

The following example will illustrate how the ϕ values given in Table 1, and a set of t-tables, are used for tests involving the unilateral and bilateral statistics. Suppose there are two Normal populations as described in §1, but it is known *a priori* that $\sigma_1 > \sigma_2$. It is desired to test H_0: $\mu_1 = \mu_2$ at the 5 % level, and a unilateral statistic would be appropriate.

A sample of size 5 is drawn from the first population and of size 9 from the second

$$(n_1 = 5,\ n_2 = 9).$$

The sample means and sums of squares of deviations are as follows:

$$\bar{x}_1 = 2{\cdot}9, \quad \Sigma_1 = 15{\cdot}0; \quad \bar{x}_2 = 0{\cdot}2, \quad \Sigma_2 = 30{\cdot}0.$$

Using a unilateral statistic it is found from (22) and (23) that

$$r_1 = \frac{(2{\cdot}7764)^2}{20} = 0{\cdot}38542, \quad r_2 = \frac{2{\cdot}630}{72} = 0{\cdot}03653.$$

The value of the statistic is

$$\frac{(2{\cdot}7)^2}{(0{\cdot}38542)\,(15{\cdot}0) + (0{\cdot}03653)\,(30{\cdot}0)} = 1{\cdot}06$$

and H_0 is rejected.

If we were to ignore the information that $\sigma_1 > \sigma_2$ and use a bilateral statistic the values of the constants would, from (26), be

$$r_1 = 0{\cdot}38542 \quad \text{and} \quad r_2 = 0{\cdot}073856.$$

The value of the statistic is

$$\frac{(2{\cdot}7)^2}{(0{\cdot}38542)(15{\cdot}0)+(0{\cdot}073856)(30{\cdot}0)} = 0{\cdot}91$$

and H_0 is not rejected.

It is noted that if the information that $R > 1$ is ignored, it is possible to arrive at a different decision than if the information is used.

Table 1. *Constants for unilateral statistics*

D.F.		$\phi = n_2 f_2 r_2$	
f_1	f_2	Test size 0·05	Test size 0·01
2	2	2·327	2·682
2	4	1·644	1·836
2	6	1·444	1·597
2	8	1·395	1·477
4	2	5·468	9·600
4	4	3·337	4·886
4	6	2·849	3·953
4	8	2·630	2·630
6	2	7·668	17·976
6	4	4·304	7·360
6	6	3·607	5·733
6	8	3·309	5·068
8	2	9·222	26·040
8	4	4·910	9·374
8	6	4·029	7·568
8	8	3·719	6·089

Table 2. *Values of size for unilateral statistics*

D.F.		$\alpha = 0{\cdot}05$					$\alpha = 0{\cdot}01$				
f_1	f_2	$R=1$	$R=4$	$R=10$	$R=100$	$R=\infty$	$R=1$	$R=4$	$R=10$	$R=100$	$R=\infty$
2	2	0·05	0·0450	0·0465	0·0494	0·05	0·01	0·0088	0·0092	0·0099	0·01
2	4	·05	·0479	·0487	·0498	·05	0·1	·0095	·0097	·0100	·01
2	6	·05	·0488	·0493	·0499	·05	·01	·0098	·0098	·0100	·01
2	8	·05	·0491	·0495	·0499	·05	·01	·0098	·0099	·0100	·01
4	2	0·05	0·0389	0·0407	0·0482	0·05	0·01	0·0065	0·0070	0·0093	0·01
4	4	·05	·0453	·0469	·0495	·05	·01	·0085	·0090	·0098	·01
4	6	·05	·0474	·0484	·0498	·05	·01	·0092	·0095	·0099	·01
4	8	·05	·0483	·0490	·0499	·05	·01	·0094	·0097	·0100	·01
6	2	0·05	0·0357	0·0364	0·0466	0·05	0·01	0·0052	0·0051	0·0084	0·01
6	4	·05	·0433	·0451	·0492	·05	·01	·0076	·0081	·0097	·01
6	6	·05	·0461	·0475	·0496	·05	·01	·0086	·0091	·0099	·01
6	8	·05	·0475	·0485	·0498	·05	·01	·0091	·0094	·0099	·01
8	2	0·05	0·0342	0·0335	0·0451	0·05	0·01	0·0046	0·0041	0·0074	0·01
8	4	·05	·0412	·0435	·0489	·05	·01	·0068	·0074	·0095	·01
8	6	·05	·0454	·0467	·0495	·05	·01	·0076	·0084	·0097	·01
8	8	·05	·0468	·0479	·0497	·05	·01	·0087	·0092	·0099	·01

Table 3. *Power of tests using unilateral statistics* ($\alpha = 0{\cdot}05$)

D.F. f_1	D.F. f_2		$R = 1$			$R = 4$			$R = 10$		
2	2	(*a*)	1·53	2·45	4·00	2·89	4·83	7·07	4·69	7·66	11·33
		(*b*)	·27	·54	·88	·28	·58	·84	·30	·59	·85
		(*c*)	·30	·62	·95	·40	·80	·98	·47	·84	·99
2	4		1·50	2·83	4·00	2·82	4·95	7·33	4·68	7·75	11·89
			·29	·68	·89	·28	·61	·86	·31	·60	·88
			·41	·90	·99	·47	·91	1·00	·55	·93	1·00
2	6		1·51	2·67	4·10	2·85	5·43	7·68	4·72	8·34	11·79
			·30	·65	·91	·29	·67	·89	·31	·65	·87
			·45	·96	1·00	·54	·97	1·00	·60	·97	1·00
2	8		1·65	2·83	4·32	2·94	5·76	7·60	4·73	8·90	11·74
			·31	·63	·89	·31	·72	·88	·31	·70	·87
			·61	·97	1·00	·60	·99	1·00	·63	·99	1·00
4	2		1·16	1·93	3·27	1·99	3·19	4·76	3·24	4·58	6·83
			·26	·59	·95	·30	·65	·94	·37	·63	·92
			·27	·60	·96	·35	·71	·96	·43	·71	·96
4	4		1·02	1·79	2·83	1·90	3·16	4·47	3·18	4·45	6·63
			·29	·69	·97	·33	·71	·94	·39	·64	·92
			·30	·70	·97	·39	·80	·97	·47	·75	·97
4	6		0·98	1·55	2·62	1·89	2·57	4·34	3·20	3·87	6·55
			·31	·62	·96	·34	·55	·93	·40	·53	·92
			·34	·66	·97	·43	·66	·98	·51	·36	·98
4	8		0·97	1·58	2·49	1·91	2·86	4·29	2·91	4·59	6·50
			·31	·67	·96	·35	·65	·93	·34	·68	·92
			·36	·74	·98	·45	·78	·98	·45	·82	·98
6	2		1·20	1·83	2·68	1·65	2·52	3·68	2·30	3·51	5·14
			·29	·56	·84	·27	·56	·88	·28	·57	·89
			·34	·64	·92	·34	·64	·92	·34	·64	·92
6	4		0·83	1·37	2·03	1·39	2·24	3·17	2·21	3·50	4·78
			·25	·56	·87	·27	·59	·88	·31	·64	·89
			·25	·56	·88	·30	·63	·90	·36	·69	·92
6	6		0·72	1·31	1·85	1·34	2·24	2·93	2·17	3·43	4·34
			·23	·61	·89	·28	·63	·85	·31	·64	·83
			·24	·61	·89	·31	·67	·89	·36	·70	·89
6	8		0·68	1·24	1·75	1·31	2·26	2·86	2·15	3·51	4·30
			·23	·61	·88	·28	·66	·85	·31	·67	·83
			·24	·62	·90	·32	·71	·90	·37	·75	·90
8	2		0·94	1·76	2·75	1·53	2·33	3·42	2·08	3·18	4·66
			·20	·52	·84	·26	·55	·86	·28	·58	·89
			·25	·66	·96	·36	·66	·94	·36	·66	·94
8	4		0·79	1·37	2·16	1·39	1·97	3·11	1·98	2·92	4·43
			·25	·60	·93	·33	·58	·93	·33	·61	·93
			·26	·60	·94	·36	·60	·94	·36	·64	·94
8	6		0·71	1·13	1·89	1·21	1·88	2·87	1·77	2·96	4·19
			·26	·55	·94	·29	·60	·92	·29	·65	·91
			·26	·55	·94	·32	·62	·94	·32	·68	·94
8	8		0·67	1·05	1·76	1·18	1·83	2·79	1·75	2·93	4·14
			·26	·56	·94	·29	·60	·92	·29	·65	·91
			·74	·56	·94	·32	·63	·94	·32	·69	·94

(*a*) Value of δ/σ_2.
(*b*) Power of test using appropriate unilateral statistic.
(*c*) Power of corresponding test using u_R^2.

References

Aspin, A. A. (1949). Tables for use in comparisons whose accuracy involves two variances, separately estimated. *Biometrika*, **36**, 290–6.

Box, G. E. P. (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems. I. Effect of inequality of variance in the one-way classification. *Ann. Math. Statist.* **25**, 290–302.

Chand, Uttam (1950). Distributions related to comparison of two means and two regression coefficients. *Ann. Math. Statist.* **21**, 507–22.

Gronow, G. D. C. (1951). Test for the significance of the difference between means in two normal populations having unequal variances. *Biometrika*, **38**, 252–6.

Hsu, P. L. (1938). Contributions to the theory of Student's *t*-test as applied to the problem of two samples. *Stat. Res. Mem.* **2**, 1–24.

Pearson, Karl, Editor (1934). *Tables of the Incomplete Beta Function*. Cambridge University Press.

Tang, P. C. (1938). The power function of the analysis of variance tests with tables and illustrations of their use. *Stat. Res. Mem.* **2**, 126–49.

Ura, Shoji (1955). On the power function of Welch's test procedure in the two sample problem. *Rep. Stat. Appl. Res., J.U.S.E.*, **4**, 1–13.

Wald, A. (1955). Testing the difference between the means of two normal populations with unknown standard deviations, from *Selected Papers in Statistics and Probability*, by Abraham Wald. New York: McGraw-Hill.

Welch, B. L. (1937). The significance of the difference between two means when the population variances are unequal. *Biometrika*, **29**, 350–62.

Welch, B. L. (1947). The generalization of Student's problem when several different population variances are involved. *Biometrika*, **34**, 28–35.

Biometrika (1960), **47**, 3 *and* 4, *p.* 355
Printed in Great Britain

Power of some two-sample non-parametric tests

BY BALKRISHNA V. SUKHATME

Michigan State University and Institute of Agricultural Research Statistics, New Delhi

1. INTRODUCTION

The problem of two samples has been considered by several authors from various points of view. Most of this work is devoted to testing hypotheses concerning differences in location of two populations. A large number of non-parametric tests have been suggested. Particular mention may be made of the Wilcoxon test which has been considered by several authors especially Mann & Whitney (1947), Lehmann (1951), Sundrum (1953) and Mood (1954). Hodges & Lehmann (1956) have shown that the Wilcoxon test is a good competitor to the Student's t-test. However, very little has been done in the direction of testing differences in the dispersion of two populations. Recently, attempts have been made by several authors, especially Mood (1954), Lehmann (1951), Sukhatme (1957, 1958*a*), Rosenbaum (1953) and Kamat (1956) to suggest suitable non-parametric tests for testing differences in dispersion. The tests proposed by the first three authors are based on what are known as generalized U statistics and the present author has recently examined the relative efficiency of these tests. However, nothing is known concerning the power of the tests suggested by the last two authors. This paper is a contribution in that direction and discusses the power of these two tests.

2. WILKS–ROSENBAUM TEST

The test is based on the number of observations of one sample, say the X's, which lie outside the extreme values of the other sample. We reject the hypothesis if this number is too large. Let $X_1, X_2, \ldots, X_m$ and $Y_1, Y_2, \ldots, Y_n$ be two samples of independent observations drawn from populations with cumulative distribution functions F and G and density functions f and g, respectively. In symbols, the test statistic may then be defined as

$$W = m - \sum_{j=1}^{m} H(Y^{(1)}, X_j, Y^{(n)}), \tag{2·1}$$

where $Y^{(1)}$ and $Y^{(n)}$ are the lowest and highest observations among the Y's in order of magnitude and

$$H(u, v, w) = 1 \quad \text{if} \quad u \leqslant v \leqslant w, \tag{2·2}$$
$$= 0, \text{ otherwise.}$$

Distribution of W. It is easy to see that

$$P(W = r) = n(n-1)\binom{m}{r}\int_{u<v} [F(v)-F(u)]^{m-r}[1+F(u)-F(v)]^{r} \times [G(v)-G(u)]^{n-2}\,dG(u)\,dG(v) \tag{2·3}$$

for $r = 0, 1, \ldots, m.$

Putting $F = G$, we obtain

$$P(W = r) = n(n-1)\binom{m}{r} B(m+n-r-1,\ r+2)$$

which is the result given by Rosenbaum (1953) and Wilks (1942). To obtain the moments of the distribution, we have

$$E(W) = m - mP[Y^{(1)} < X_j < Y^{(n)}]$$

$$= m - mn(n-1)\int_{u<v} [F(v)-F(u)]\,[G(v)-G(u)]^{n-2}\,dG(u)\,dG(v). \tag{2·4}$$

$$\text{var } W = \text{var} \sum_{j=1}^{m} H(Y^{(1)}, X_j, Y^{(n)})$$

$$= E\sum_j H(Y^{(1)}, X_j, Y^{(n)})$$

$$+ E\sum_{j\neq k} H(Y^{(1)}, X_j, Y^{(n)})\,H(Y^{(1)}, X_k, Y^{(n)})$$

$$- [E\sum_j H(Y^{(1)}, X_j, Y^{(n)})]^2. \tag{2·5}$$

Now $E\sum_{j\neq k} H(Y^{(1)}, X_j, Y^{(n)})\,H(Y^{(1)}, X_k, Y^{(n)})$

$$= m(m-1)\,n(n-1)\int_{u<v} [F(v)-F(u)]^2\,[G(v)-G(u)]^{n-2}\,dG(v)\,dG(u),$$

whence

$$\text{var } W = mn(n-1)\int_{u<v} [F(v)-F(u)]\,[G(v)-G(u)]^{n-2}\,dG(u)\,dG(v)$$

$$+ m(m-1)\,n(n-1)\int_{u<v} [F(v)-F(u)]^2\,[G(v)-G(u)]^{n-2}\,dG(u)\,dG(v)$$

$$- m^2n^2(n-1)^2\left[\int_{u<v} [F(v)-F(u)]\,[G(v)-G(u)]^{n-2}\,dG(u)\,dG(v)\right]^2. \tag{2·6}$$

Substituting $F = G$, we obtain the moments under the hypothesis tested.

3. KAMAT'S TEST

The test statistic is defined as

$$D = R_m - R_n + n, \tag{3·1}$$

where R_m and R_n are the ranges of the ranks of the X's and the Y's in the combined sample, respectively. We reject the hypothesis if D is either too large or too small.

Distribution of D under the alternative. In what follows, we assume that X's and Y's are ordered so that

$$X_1 < X_2 < \ldots < X_m,$$
$$Y_1 < Y_2 < \ldots < Y_n.$$

As shown by Kamat, the totality of values of D can be built up in four distinct ways:

(i) $R_m = m-1$, $R_n = n-1$; $D = m$.

This can happen in only two ways. The X's should occupy either the first m places or the last m. Clearly,

$$P(D = m) = P(Y_n < X_1) + P(X_m < Y_1)$$

$$= n\int F^m[1-G]^{n-1}\,dG + m\int G^n[1-F]^m\,dF. \tag{3·2}$$

(ii) $R_m = m-1+i$, $R_n = n-1+m$, $D = i$ $(i = 0, 1, \ldots, n-1)$.

Then $$P(D=i)=\sum_{s=1}^{n-i-1} P(Y_s<X_1<Y_{s+1}<\ldots<Y_{s+i}<X_m<Y_{s+i+1})$$

$$=\frac{n!\,m!}{(m-2)!\,i!\,(n-1)!}\int_{u<v}\{[1+G(u)-G(v)]^{n-i}-[G(u)]^{n-i}-[1-G(v)]^{n-i}\}$$

$$\times[G(v)-G(u)]^{i}[F(v)-F(u)]^{m-2}\,dF(u)\,dF(v). \quad (3{\cdot}3)$$

(iii) $R_m=m-1+n$, $R_n=n-1+j$, $D=n+m-j$ $(j=0,1,\ldots,m-1)$.
By symmetry, as in (ii)

$$P(D=m+n-j)=\frac{n!\,m!}{(n-2)!\,(m-j)!\,j!}\int_{u<v}\{[1+F(u)-F(v)]^{m-j}-[1-F(v)]^{m-j}-[F(u)]^{m-j}\}$$

$$\times[F(v)-F(u)]^{j}[G(v)-G(u)]^{n-2}\,dG(u)\,dG(v). \quad (3{\cdot}4)$$

(iv) $R_m=m-1+i$, $R_n=n-1+j$, $D=m+i-j$.
This can happen either when the sequence starts with X_1 or Y_1.

$$P(D=m+i-j)$$

$$=(X_{m-j}<Y_1<(j-1)\,X\text{'s} \quad \text{and} \quad (i-1)\,Y\text{'s}<X_m<Y_{i+1})$$

$$+P(Y_{n-i}<X_1<(j-1)\,X\text{'s} \quad \text{and} \quad (i-1)\,Y\text{'s}<Y_m<X_{j+1})$$

$$=\frac{m!}{(m-j)\,(j-1)!}\,\frac{n!}{(n-i)\,(i-1)!}\,[A_1+A_2], \quad (3{\cdot}5)$$

where $$A_1=\int_{u<v}[F(u)]^{m-j}[F(v)-F(u)]^{j-1}[G(v)-G(u)]^{i-1}[1-G(v)]^{n-i}\,dG(u)\,dF(v),$$

$$A_2=\int_{u<v}G(u)^{n-i}[F(v)-F(u)]^{j-1}[G(v)-G(u)]^{i-1}[1-F(v)]^{m-j}\,dG(v)\,dF(u).$$

Putting $F=G$ in (i), (ii), (iii) and (iv), we obtain the distribution of D under the hypothesis as given by Kamat.

To obtain the mean and the variance of D, we express the statistic D in the form

$$D=(s_m-s_1)-(r_n-r_1)+n,$$

where s_i is the rank of X_i and r_j, the rank of Y_j in the combined sample of X's and Y's.

Then $$s_i=i+\sum_{k=1}^{n}\phi(Y_k,X_i)$$

and $$r_j=j+\sum_{t=1}^{m}\phi(X_t,Y_j),$$

where $$\phi(u,v)=1 \quad \text{if} \quad u\leqslant v,$$

$$=0, \text{ otherwise.}$$

$$E(D)=E(s_m-s_1)-E(r_n-r_1)+n,$$

$$\operatorname{var}D=\operatorname{var}(s_m-s_1)+\operatorname{var}(r_n-r_1)-2\cos(s_m-s_1,r_n-r_1).$$

Since $$r_n-r_1=(n-1)+\sum_{j=1}^{m}H(Y_1,X_j,Y_n),$$

$$E(r_n-r_1)=(n-1)+mn(n-1)\int_{u<v}[F(v)-F(u)]\,[G(v)-G(u)]^{n-2}\,dG(u)\,dG(v),$$

$$\operatorname{var}(r_n-r_1)=\operatorname{var}\sum_{j=1}^{m}H(Y_1,X_j,Y_n)=\operatorname{var}W.$$

Similarly, we have

$$E(s_m - s_1) = (m-1) + nm(m-1)\int_{u<v} [G(v)-G(u)]\,[F(v)-F(u)]^{m-2}\,dF(u)\,dF(v),$$

$$\begin{aligned}\operatorname{var}(s_m - s_1) = {} & nm(m-1)\int_{u<v} [G(v)-G(u)]\,[F(v)-F(u)]^{m-2}\,dF(u)\,dF(v) \\ & + m(m-1)\,n(n-1)\int_{u<v} [G(v)-G(u)]^2\,[F(v)-F(u)]^{m-2}\,dF(u)\,dF(v) \\ & - n^2m^2(m-1)^2\left[\int_{u<v} [G(v)-G(u)]\,[F(v)-F(u)]^{m-2}\,dF(v)\,dF(u)\right]^2,\end{aligned}$$

$$\operatorname{cov}(s_m - s_1, r_n - r_1) = Es_m r_n - Es_m r_1 - Es_1 r_n + Er_1 s_1 - E(s_m - s_1)\,E(r_n - r_1).$$

It can be shown that

$$E(r_1 s_1) = 1 + mn\int F[1-G]^{n-1}\,dG + mn\int G[1-F]^{m-1}\,dF.$$

$$\begin{aligned}E(r_1 s_m) = {} & m + m^2n\int F[1-G]^{n-1}\,dG + mn\int GF^{m-1}\,dF \\ & + nm(m-1)\int_{u<v} [F(v)]^{m-2}\,[1-G(u)]^{n-1}\,F(u)\,dG(u)\,dF(v) \\ & + m(m-1)\,n(n-1)\int_{u<v} [F(v)]^{m-2}\,[1-G(u)]^{n-2}\,F(u)\,[G(v)-G(u)]\,dG(u)\,dF(v).\end{aligned}$$

$$\begin{aligned}E(s_1 r_n) = {} & n + mn\int FG^{n-1}\,dG + n^2m\int G[1-F]^{m-1}\,dF \\ & + mn(n-1)\int_{u<v} [1-F(u)]^{m-1}\,[G(v)]^{n-2}\,G(u)\,dF(u)\,dG(v) \\ & + m(m-1)\,n(n-1)\int_{u<v} [1-F(u)]^{m-2}\,[G(v)]^{n-2}\,G(u)\,[F(v)-F(u)]\,dF(u)\,dG(v).\end{aligned}$$

$$\begin{aligned}E(r_n s_m) = {} & mn + m^2n\int FG^{n-1}\,dG + n^2m\int GF^{m-1}\,dF \\ & + mn\int [F-F^m]\,G^{n-1}\,dG + mn\int [G-G^n]\,F^{m-1}\,dF \\ & + m(m-1)\,n(n-1)\int_{u<v} [F(u)]^{m-1}\,G(u)\,[G(v)]^{n-2}\,dF(u)\,dG(v) \\ & + m(m-1)\,n(n-1)\int_{u<v} [G(u)]^{n-1}\,F(u)\,[F(v)]^{m-2}\,dG(u)\,dF(v).\end{aligned}$$

Using these results, we obtain the mean and the variance of D under the alternative. Substituting $F = G$, we obtain

$$E(s_m - s_1) = \frac{(m-1)(n+m+1)}{m+1},$$

$$E(r_n - r_1) = \frac{(n-1)(m+n+1)}{n+1},$$

$$\operatorname{var}(s_m - s_1) = \frac{2n(m-1)(m+n+1)}{(m+1)^2(m+2)},$$

$$\operatorname{var}(r_n - r_1) = \frac{2m(n-1)(m+n+1)}{(n+1)^2(n+2)},$$

$$\operatorname{cov}(r_n - r_1, s_m - s_1) = -\frac{4mn}{(m+1)(n+1)} + 2 - \frac{2}{\binom{m+n}{n}}.$$

Whence we obtain the mean and variance of D under the hypothesis as given by Kamat.

4. Power of the two tests

In this section we will consider the power of these two tests for small samples and for different levels of significance. For the sake of simplicity we will consider one sided alternatives only.

Let $m = n = 5$ and $\alpha \leqslant 0{\cdot}05$. From the tables given by Rosenbaum and Kamat, the 5 % points of W and D are 5 and 10, respectively, so that

$$P(W \geqslant 5) \leqslant 0{\cdot}05, \quad P(D \geqslant 10) \leqslant 0{\cdot}05.$$

Hence the power of the W test is given by

$$\begin{aligned}\beta_W &= P(W \geqslant 5 \,|\, A) \\ &= 20 \int_{u<v} [1 + F(u) - F(v)]^5 [G(v) - G(u)]^3 \, dG(u) \, dG(v).\end{aligned}$$

The power of the D test is given by

$$\begin{aligned}\beta_D &= P(D \geqslant 10 \,|\, A) \\ &= 20 \int_{u<v} \{[1 + F(u) - F(v)]^5 - [F(u)]^5 - [1 - F(v)]^5\} [G(v) - G(u)]^3 \, dG(u) \, dG(v).\end{aligned}$$

Clearly $\beta_W > \beta_D$ for all alternatives.

The same procedure was carried out in a few more cases and it was found that in all these cases the power of the D test is smaller than the power of the W test. These cases are listed in Table 1.

Table 1

Sample size $m = n$	5	5	6	6	7	8
Level of significance	0·025	0·05	0·01	0·05	0·01	0·01

From the above results, we thus see that at least for small samples and low level of significance the D test is less powerful than the W test. As the calculations involved are particularly tedious, detailed investigations in regard to power were made for normal alternatives for the cases $m = n = 5$ and $m = n = 6$ only. Let $F(x) = G(x\theta)$, θ being a scale parameter and $G(x)$ the normal distribution function. Using the methods of numerical integration, power was computed for different alternatives $\theta = 0{\cdot}9$, $0{\cdot}8$, $0{\cdot}4$ and $0{\cdot}1$ for different levels of significance. These results are given in Tables 2 and 3. The exact size of the test is also given in each case.

Table 2. $m = n = 5$

	Test	Exact size of the test	Power of the test against the alternative $\theta = 0{\cdot}9$	$\theta = 0{\cdot}8$	$\theta = 0{\cdot}4$	$\theta = 0{\cdot}1$
$\alpha \leqslant 0{\cdot}025$	W	0·0238	0·0285	0·0331	0·1538	0·5526
	D	·0159	·0206	·0252	·1396	·5144
$\alpha \leqslant 0{\cdot}11$	W	·1032	·1106	·1180	·4361	·8119
	D	·0635	·0695	·0755	·3728	·9622

Table 3. $m = n = 6$

	Test	Exact size of the test	Power of the test against the alternative $\theta = 0{\cdot}9$	$\theta = 0{\cdot}8$	$\theta = 0{\cdot}4$	$\theta = 0{\cdot}1$
$\alpha \leqslant 0{\cdot}01$	W	0·0076	0·0095	0·0115	0·0873	0·4644
	D	·0054	·0074	·0093	·0827	·4478
$\alpha \leqslant 0{\cdot}05$	W	·0400	·0498	·0595	·2945	·7605
	D	·0271	·0362	·0453	·2692	·8424

It will be seen from Tables 2 and 3 that the power of the W test is in general larger than that of the D test, but that the exact size of the W test is always larger than that of the D test. This makes it difficult to compare the two tests. When the exact size of each test was calculated for different levels of significance and different sample sizes, the results given in Table 4 were obtained. It will be seen that the exact size of the W test is in general larger than that of the D test at least for low levels of significance.

Table 4. *Exact size of the W and D tests*

Sample size $m = n$	$\alpha = 0{\cdot}01$ W	D	$\alpha = 0{\cdot}025$ W	D	$\alpha = 0{\cdot}05$ W	D
5	—	—	0·0238	0·0159	0·0238	0·0159
6	0·0067	0·0054	·0076	·0054	·0400	·0271
7	·0023	·0020	·0146	·0120	·0146	·0406
8	·0050	·0038	·0203	·0148	·0203	·0422
9	·0076	·0058	·0249	·0185	·0249	·0475
10	·0099	·0076	·0099	·0216	·0286	·0216

Table 5. $m = n = 5$

	Test	Power of the test against the alternative $\theta = 0{\cdot}9$	$\theta = 0{\cdot}8$	$\theta = 0{\cdot}4$	$\theta = 0{\cdot}1$
$\alpha = 0{\cdot}025$	W	0·0297	0·0344	0·1581	0·5565
	D	·0299	·0349	·1842	·6000
	F	·0336	·0533	·3520	·9772
$\alpha = 0{\cdot}05$	W	·0556	·0611	·2470	·6382
	D	·0556	·0613	·3068	·8354
	F	·0686	·1040	·5000	·9896

Table 6. $m = n = 6$

	Test	Power of the test against the alternative $\theta = 0{\cdot}9$	$\theta = 0{\cdot}8$	$\theta = 0{\cdot}4$	$\theta = 0{\cdot}1$
$\alpha = 0{\cdot}025$	W	0·0311	0·0373	0·2468	0·6924
	D	·0335	·0419	·2515	·8049
	F	·0400	·0608	·4492	·9935
$\alpha = 0{\cdot}01$	W	·0125	·0151	·1027	·4865
	D	·0135	·0169	·1223	·5315
	F	·0154	·0237	·2715	·9846

To make a fairer comparison possible, the two tests were randomized so that each had the same level of significance. For example, in the case $m = n = 5$ for neighbouring integer critical values of W, the W test has exact levels $\alpha = 0{\cdot}0238$ and $\alpha = 0{\cdot}1032$. If we use the former point with chance 391/397 and the latter with chance 6/397, an effective significance level of $\alpha = 0{\cdot}025$ will result. These values of the power based on tests adjusted in this way are given in Tables 5 and 6. It will be observed that now the power of the D test is always greater than that of the W test, although the difference between the powers for the two tests is naturally very small for alternatives close to the hypothesis.

We shall now compare these two non-parametric tests with the most powerful, variance-ratio F test. From the results given in Tables 5 and 6, it is clear that under conditions where the F test is applicable the power of the two non-parametric tests is much less than that of the F test. Suppose that a test T_1, of level of significance α, and using N_1 observations, has power $\beta_{T_1}(N_1, \alpha, \theta)$ against alternative θ. If a test T_2, also of level α, requires N_2 observations to produce the same power for the same alternative, the efficiency of T_2 with respect to T_1, may be defined as the ratio N_2/N_1. Table 7 gives the efficiency for each of the two tests with respect to the variance ratio F test for normal alternatives. It will be seen that the D test is always more efficient than the W test.

Table 7

Sample size	Level of significance	Test	Efficiency of the test against the alternative $\theta = 0{\cdot}9$	$\theta = 0{\cdot}4$	$\theta = 0{\cdot}1$
$m = n = 5$	$\alpha = 0{\cdot}025$	D	0·59	0·69	0·55
		W	·57	·64	·53
	$\alpha = 0{\cdot}05$	D	·40	·69	·60
		W	·40	·60	·50
$m = n = 6$	$\alpha = 0{\cdot}01$	D	·67	·67	·52
		W	·62	·61	·49
	$\alpha = 0{\cdot}025$	D	·58	·68	·57
		W	·52	·68	·49

Summary

Rosenbaum (1953) and Kamat (1956) have proposed non-parametric tests for testing the hypothesis that two samples come from the same population against the alternative that the Y's are more spread out than the X's and vice versa. This paper gives the exact distribution of the two test statistics under the alternative hypothesis and a discussion concerning the power of these two tests against scalar alternatives for small samples.

I wish to thank Prof. E. S. Pearson for his useful comments.

References

Hodges, J. L. & Lehmann, E. L. (1956). The efficiency of some non-parametric competitors of the t test. *Ann. Math. Statist.* **27**, 324.

Kamat, A. R. (1956). A two sample distribution free test. *Biometrika.* **43**, 377.

Lehmann, E. L. (1951). Consistency and unbiasedness of certain non-parametric tests. *Ann. Math. Statist.* **22**, 165.

Mann, H. B. & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *Ann. Math. Statist.* **18**, 50.

Mood, A. M. (1954). On the asymptotic efficiency of certain non-parametric two sample tests. *Ann. Math. Statist.* **25**, 514.

ROSENBAUM, S. (1953). Tables for a non-parametric test for dispersion. *Ann. Math. Statist.* **24**, 663.
SUKHATME, B. V. (1957). On certain two sample non-parametric tests for comparing variances. *Ann. Math. Statist.* **28**, 188.
SUKHATME, B. V. (1958*a*). A two sample distribution free test for comparing variances. *Biometrika*, **45**, 544.
SUKHATME, B. V. (1958*b*). Testing the hypothesis that two populations differ only in location. *Ann. Math. Statist.* **29**, 60.
SUNDRUM, R. M. (1953). Power of the Wilcoxon two sample test. *J.R. Statist. Soc.* B, **15**, 248.
WILCOXON, F. (1954). Individual comparison by ranking method. *Biometrics Bulletin*, **1**, 80.
WILKS, S. S. (1942). Statistical prediction with special reference to the problem of tolerance limits. *Ann. Math. Statist.* **13**, 400.

Biometrika (1960), **47**, 3 *and*, 4 *p.* 363
With 7 text-figures
Printed in Great Britain

Sampling inspection of continuous processes with no autocorrelation between successive results

BY W. D. EWAN AND K. W. KEMP
British Nylon Spinners Ltd., Pontypool, Monmouthshire

1. INTRODUCTION

We consider here the problem of controlling the quality of the output of a continuous manufacturing process where quality will normally be maintained at an acceptable level for long periods of time, but where a change can take place which will result in the process manufacturing material to a quality level which is unacceptable.

Our first attempts in this field were to take standard sequential and multiple sampling schemes derived for batch inspection problems and to use these schemes retrospectively, the first sample being the current sample the second sample being that taken immediately previous to the current sample and so on. An example of such a scheme is quoted by Barnard (1954).

Page (1954) introduced the concept of cumulative sum charts based on Wald sequential schemes and since this original paper we have adopted this principle to give a wide range of inspection schemes with known properties. Barnard (1959) has recently drawn attention again to the cumulative sum chart and has prepared certain empirical methods for deciding when changes have occurred and for estimating the magnitude of these.

We describe, in this paper, the methods which we have developed for our own use and which we have now been applying in practice since 1955. The method is entirely systematic, the parameters of a scheme with required properties can be determined easily, and in our experience the technique is very simple to apply either in graphical or tabular form.

2. CUMULATIVE SUM CHART

An advantage of the cumulative sum chart is that a change in quality can be seen much more easily by visual inspection than from the normal type of control chart where the results are plotted independently of each other. Fig. 1 shows a control chart where points on the first part of the chart have been sampled from a normal population with zero mean and unit standard deviation: half way along the figure the population mean has been changed to 0·20. Fig. 2 shows the cumulative sum chart of these results as in the presentation by Barnard (1959). It is much easier to detect the change from visual inspection of Fig. 2 than from Fig. 1. The magnitude of a change can be determined from the slope of the cumulative sum chart and both Page (1957) and Barnard (1959) have considered this question and that of determining the point at which a change occurs.

If k is subtracted from each sample value and the resulting figure plotted the cumulative chart will slope downwards if the population mean m is less than k and upwards if it is greater than k. Schemes for which the cumulative chart slopes downwards for quality which is acceptable and upwards when quality is rejectable are similar to those described by Page (1954). We shall call k the *reference value*. Page suggests that the decision that quality has

changed should be made after the cumulative sum has changed direction and has exceeded the previous lowest result by a value of h or more. We shall call h the *decision interval.*

Fig. 3 shows a comparison between the results of Fig. 1 and the same results plotted on a cumulative sum chart with lower boundary zero and upper boundary h using a reference value $k = 0{\cdot}10$. Whenever a result greater than 0·10 is obtained the sum of the deviations from this value is accumulated until the sum returns below zero or exceeds the decision

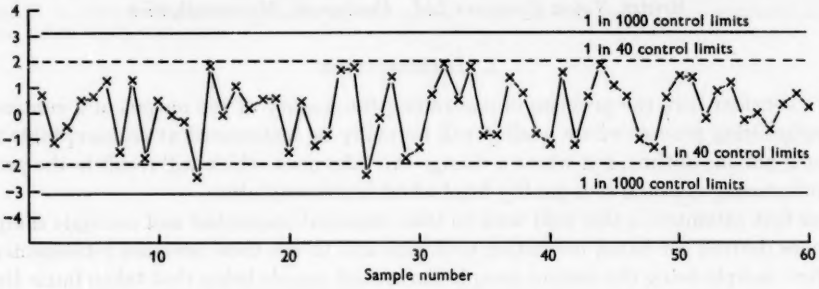


Fig. 1. Standard control chart. Mean of first thirty results = 0·00.
Mean of second thirty results = 0·20.

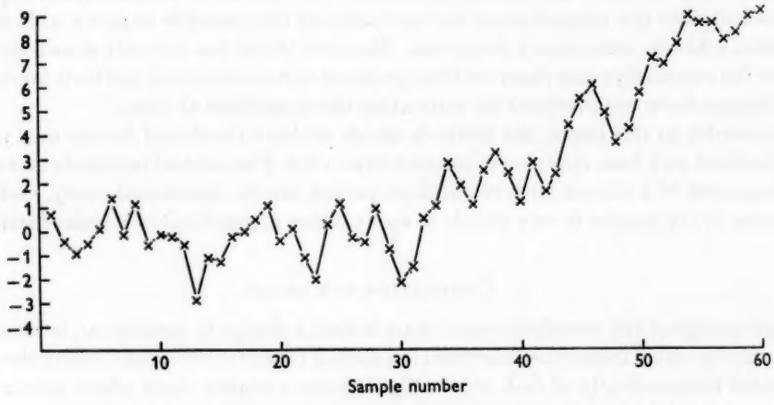


Fig. 2. Cumulative sum chart. Mean of first thirty results = 0·00.
Mean of second thirty results = 0·20.

interval. This type of chart is effectively very similar to that in Fig. 2 and a decision that quality has changed can be reached at the same time with both. A practical advantage of the method demonstrated in Fig. 3 is that the chart is bounded in the sense that the cumulative sum does not run off the paper. The principal involved is easily applicable to a table of successive results without recourse to plotting if no graph is desired. Fig. 2 provides a useful record of the past history of *all* the results that have been obtained, whilst the chart of Fig. 3 only records results which are likely to be relevant to the detection of a deterioration of quality.

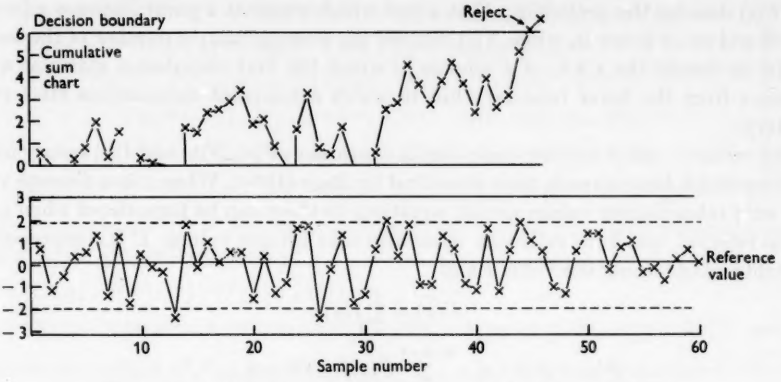


Fig. 3. Control chart and cumulative sum plotted when required.

3. Average run length and average duration of an inspection scheme

In order to present the properties of a continuous sampling inspection scheme we shall make use of the concept of the average run length (A.R.L.) described by Page (1954). The definition of the A.R.L. used by us is the same as that given by Barnard and is slightly different from that used by Page. We define the *average run length* as follows: when the quality remains constant the average run length of an inspection scheme is the expected number of *samples* obtained before action is taken.* An important property of any continuous sampling scheme is the average amount of material produced at a given quality level before taking action. For a continuous manufacturing process producing at a constant rate this is directly related to the time which elapses before action is taken. We can define this quantity as the *average duration* of the inspection scheme. After a change in quality has occurred the average duration of a scheme using the cumulative sum chart will depend upon the position of the cumulative sum at the time when the first sample after the change is taken. If the process is sampled every s hours (say) and the sample is taken from current production an approximate estimate of the average duration of a scheme can be obtained by using the simple expression $(\text{A.R.L.} - 1)s + t$. Here t is the average time which elapses between the moment that the change occurred and the time that the first sample was taken. If the probability of the process changing to rejectable quality level (R.Q.L.) is constant with respect to time and is small relative to the sampling interval, t is approximately $\frac{1}{2}s$ and the average duration of the scheme at this quality is $(\text{A.R.L.} - \frac{1}{2})s$.

Although the A.R.L. is used in this paper to characterize the sampling schemes discussed it is obvious that other parameters can be devised. For example, the frequency of obtaining run lengths greater than a given value at R.Q.L. or less than another value at A.Q.L. could be more relevant in some situations. The distribution of run length is undoubtedly relevant to these and a method of obtaining it is described together with two examples.

4. Calculation of the A.R.L.

The scheme illustrated by Fig. 3 is equivalent to a succession of Wald sequential schemes with horizontal boundaries distance h apart. For convenience we take the lower line as

* Page defines the A.R.L. as the expected number of *items* sampled before action is taken.

zero. $P(z)$ denotes the probability that a test which starts at a point distance z from this line will end on or below it, while $N(z)$ denotes the average sample number of the test. We use $L(z)$ to denote the A.R.L. of a scheme in which the first cumulation starts at a point distance z from the lower boundary but in which subsequent cumulations start on this boundary.

If the variate x which is being controlled is continuous $P(z)$, $N(z)$ and $L(z)$ satisfy integral equations which have already been described by Page (1954). When x is a discrete variate which only takes integer values similar equations to these can be formulated when z, h the decision interval, and k the reference value also take integer values. If $f(x)$ represents the probability of obtaining the value x and

$$F(\alpha) = \sum_{0}^{\alpha} f(x)$$

then
$$P(z) = F(k-z) + \sum_{y=1}^{y=h-1} P(y) f(y+k-z), \tag{1}$$

$$N(z) = 1 + \sum_{y=1}^{y=h-1} N(y) f(y+k-z), \tag{2}$$

$$L(z) = 1 + L(0)\, F(k-z) + \sum_{y=1}^{y=h-1} L(y) f(y+k-z). \tag{3}$$

The method described by Kemp (1958) to obtain approximate solutions of the equations for $P(z)$ and $N(z)$ when x is normally distributed can easily be generalized for use with both continuous and discrete functions. Let m be the mean of x; further:

(i) For a continuous variate, if $f(x)\,dx$ is the probability density of x, let

$$F(\alpha) = \int_{-\infty}^{\alpha} f(x)\,dx, \quad M(\alpha) = \int_{-\infty}^{\alpha} x f(x)\,dx,$$

and
$$G(\alpha) = \int_{-\infty}^{\alpha} e^{wx} f(x)\,dx,$$

where w is a real non-zero root of the equation,

$$\int_{-\infty}^{+\infty} e^{ux} f(x)\,dx = e^{uk}.$$

(ii) For a discrete variate, let

$$F(\alpha) = \sum_{0}^{\alpha} f(x), \quad M(\alpha) = \sum_{0}^{\alpha} x f(x)$$

and
$$G(\alpha) = \sum_{0}^{\alpha} e^{wx} f(x),$$

where w is a real non-zero root of the equation

$$\sum_{0}^{\infty} e^{ux} f(x) = e^{uk}.$$

By the method mentioned we obtain for a continuous variate or a scheme with a discrete variate with integer values of h and k

$$P(0) \doteqdot \frac{\begin{vmatrix} F(k) & K_2 \\ F(k-h) & K_4 \end{vmatrix}}{\begin{vmatrix} K_1 & K_2 \\ K_3 & K_4 \end{vmatrix}}, \quad N(0) \doteqdot \frac{\begin{vmatrix} \beta_1 & K_2 \\ \beta_2 & K_4 \end{vmatrix}}{\begin{vmatrix} K_1 & K_2 \\ K_3 & K_4 \end{vmatrix}}, \tag{4}$$

where
$$\left.\begin{aligned}
K_1 &= 1+\{[F(h+k)^*-F(k)]\,e^{wh}-e^{-wk}\,[G(h+k)^*-G(k)]\}/(1-e^{wh}),\\
K_2 &= \{e^{-wk}\,[G(h+k)^*-G(k)]-[F(h+k)^*-F(k)]\}/(1-e^{wh}),\\
K_3 &= \{F(k)^*-F(k-h)-e^{-wk}\,[G(k)^*-G(k-h)]\}\,e^{wh}/(1-e^{wh}),\\
K_4 &= 1-\{F(k)^*-F(k-h)-e^{-w(k-h)}\,[G(k)^*-G(k-h)]\}/(1-e^{wh}),\\
\beta_1 &= 1-\{K_2h+M(h+k)^*-M(k)-k[F(h+k)^*-F(k)]\}/(m-k),\\
\beta_2 &= 1-\{(1-K_4)\,h+M(k)^*-M(k-h)-(k-h)\,[F(k)^*-F(k-h)]\}/(m-k).
\end{aligned}\right\} \quad (5)$$

In these equations $F(t)^* = F(t)$ for a continuous variate and $F(t-1)$ for a discrete variate; $G(t)^*$ and $M(t)^*$ are similarly defined.

$$P(z) \doteqdot \{[P(h)-P(0)\,e^{wh}]+[P(0)-P(h)]\,e^{wz}\}/(1-e^{wh}), \quad (6)$$

$$N(z) \doteqdot \{N(h)-N(0)\,e^{wh}+h/(m-k)+[N(0)-N(h)-h/(m-k)]\,e^{wz}-z\}/(m-k)\,(1-e^{wh}). \quad (7)$$

Approximate values of $P(z)$ and $N(z)$ for $z = 1, 2, \ldots, (h-1)$ can be calculated with equations (6) and (7) and these values can then be used to obtain better values of $P(0)$ and $N(0)$. A method of successive substitution can be employed which quickly leads to the true values of $P(0)$ and $N(0)$. To obtain $P(0)$, for example, when x is a discrete variate we take the linear set of equations for $P(z)$ $(z = 0, 1, 2, \ldots, h)$ generated by equation (1) and substitute the approximate values for $P(0)$, $P(1)$, ..., $P(h-1)$ into the equation for $P(h)$ and recalculate $P(h)$. This value, together with the original values of $P(0)$, $P(1)$, ..., $P(h-2)$, is then substituted in the equation for $P(h-1)$ and a new value for this is calculated and so on. The procedure is repeated until the true value of $P(0)$ is obtained. $N(0)$ can be found in a similar way. If x is a continuous variate the linear equations are formulated from the integral equations satisfied by $P(z)$ and $N(z)$ using methods of quadrature.

When $P(0)$ and $N(0)$ are known, $L(0)$ can be obtained since

$$L(0) = N(0)/[1-P(0)]. \quad (8)$$

A method which can be used for the discrete case to determine $P(0)$ and $N(0)$ exactly when the variate only assumes integer values has been described by Girshick (1946). It has been applied to the case where articles are classified as defective or non-defective and where unit samples are taken. Often in industrial practice it is neither desirable nor practical to take such samples. The method described here can be used to obtain approximate A.R.L.'s of schemes for any sample size and the exact values can be obtained if required with less computational effort than is required by the method of Girshick.

Equations (1), (2) and (3) only apply to situations where h, k and z take integer values; if these parameters can have fractional values then they must be modified. It will be convenient to postpone discussion of this particular case until we consider schemes for a Poisson variate.

5. The run-length distribution

If $p(n, z)$ is the probability that a test starting at z will have a run length n then if x is a continuous variate

$$p(n,z) = p(n-1,0)\,F(k-z)+\int_0^h p(n-1,y)f(y+k-z)\,dy. \quad (9)$$

It follows that from this equation, if $\phi(z,t)$ is the moment generating function of the run-length distribution, then

$$\phi(z,t) = \sum_1^\infty p(n,z)\,e^{nt}$$

satisfies the equation

$$e^{-t}\phi(z,t) = 1-F(h+k-z)+\phi(0,t)\,F(k-z)+\int_0^h \phi(y,t)f(y+k-z)\,dy. \tag{10}$$

If we differentiate equation (10) and put $t = 0$ in the usual way we can obtain integral equations which the moments of n satisfy. For example, after differentiating once we obtain the equation for $L(z)$ given by Page (1954), viz.

$$L(z) = 1+L(0)\,F(k-z)+\int_0^h L(y)f(y+k-z)\,dy, \tag{11}$$

whilst if $\mu_2'(z)$ denotes the second zero moment of n we find

$$\mu_2'(z) = 2L(z)-1+\mu_2'(0)\,F(k-z)+\int_0^h \mu_2'(y)f(y+k-z)\,dy. \tag{12}$$

The moment generating function of n can also be expressed as a function of the moment generating functions of the number of samples in a single acceptance test and a single rejection test. If for tests starting on the acceptance boundary these are $\phi_1(t)$ and $\phi_2(t)$, respectively, then

$$\phi(0,t) = [1-P(0)\,\phi_2(t)]/[1-P(0)\,\phi_1(t)]. \tag{13}$$

Equation (13) shows that the run-length distributions of schemes with a high average run length also have variances which are greater than the square of the A.R.L. For if $V(N)$ is the variance of the sample number of a single test and $V(n)$ is the variance of the run-length distribution, then from equation (13) we find that at acceptable quality level

$$V(n) \doteqdot L^2(0)+V(N)/[1-P(0)]. \tag{14}$$

Equation (14) shows that when $L(0)$ is large, which will generally be the case when quality is being manufactured at an acceptable level, $\sigma(n) \doteqdot L(0)$. It is therefore reasonable to anticipate that, except for very small values of n, a close approximation to $p(n,0)$ will be given by the expression

$$p(n,0) \doteqdot \frac{1}{L(0)}\exp-\frac{n-1}{L(0)}, \tag{15}$$

so that the probability of obtaining a run length $\alpha L(0)$ is

$$\sum_1^{\alpha L(0)} p(n,0) \doteqdot 1-\exp\left[\frac{1}{L(0)}-\alpha\right]. \tag{16}$$

When the probabilities of a single test ending on the upper or the lower boundary at a given stage are known it is possible to obtain the run-length distribution with use of a simple recurrence relationship. If $p_u(m)$ and $p_l(m)$ are respectively the probability that a single test will end at the mth sample on the upper or lower boundary, it can easily be shown that

$$p(n,0) = p_u(n)+\sum_{m=1}^{m=n-1} p_l(m)\,p(n-m,0). \tag{17}$$

If $g_m(y_m)$ represents the conditional frequency distribution of

$$y_m = \sum_{i=1}^{i=m} x_i$$

for a single test then

$$g_m(y_m) = \int_0^h f(y_m-y)\,g_{m-1}(y)\,dy. \tag{18}$$

$$p_u(m) = \int_h^\infty g_m(y)\,dy; \quad p_l(m) = \int_{-\infty}^0 g_m(y)\,dy. \tag{19}$$

Equations (18) and (19) are extremely useful when $p_u(m)$ and $p_l(m)$ have to be obtained by quadrature.

6. Normal variate

If x is normally distributed with mean m and unit standard deviation then

$$w = -2(m-k), \quad K_1 = K_4 \quad \text{and} \quad e^{wh}K_2 = K_3.$$

If
$$\phi(y) = (2\pi)^{-\frac{1}{2}}\exp -\tfrac{1}{2}y^2 \quad \text{and} \quad \Phi(s) = \int_{-\infty}^{s}\phi(y)\,dy,$$

we obtain
$$F(\alpha) = \Phi(\alpha-m), \quad M(\alpha) = m\Phi(\alpha-m) - (\alpha-m)\,\phi(\alpha-m),$$
$$e^{-wk}G(\alpha) = \Phi(\alpha+m-2k).$$

Approximate values for $P(0)$ and $N(0)$ can be calculated from the expressions which are obtained when these results are substituted in equations (4) and (5). More accurate values of $P(0)$ and $N(0)$ and hence $L(0)$ can then be obtained by successive substitution as already described.

A useful approximation to $L(0)$ when $m-k$ and h have the same sign is

$$L(0) \doteqdot 1 + h/(m-k). \tag{20}$$

This simple expression will give a reasonable approximation to the value of $L(0)$ provided $m-k$ is not nearly zero and provided h is sufficiently large. In fact the approximation seems to be satisfactory for values of $h \geqslant 2$. In the practical field where schemes are devised to control the quality level of continuously operating industrial processes inaccuracies incurred by the use of equation (20) over the range described will cause little concern.

7. Nomograms for the A.R.L.

Table 1. *Values of $L(0)$ for different h and $m-k$ for a normal variate with unit variance*

(*a*) Negative values of $m-k$

$h=2$		$h=3$		$h=4$		$h=5$	
$m-k$	$L(0)$	$m-k$	$L(0)$	$m-k$	$L(0)$	$m-k$	$L(0)$
−1·30	951	−0·90	1056	−0·65	954	−0·50	983
−1·25	761	−0·85	785	−0·60	668	−0·45	636
−1·15	491	−0·80	587	−0·55	481	−0·40	456
−1·00	258	−0·70	327	−0·50	341	−0·35	310
−0·80	115	−0·60	201	−0·40	181	−0·30	214
−0·60	54	−0·50	100	−0·30	103	−0·20	110

(*b*) Positive values of $m-k$

$h=2$		$h=3$		$h=4$		$h=5$	
$m-k$	$L(0)$	$m-k$	$L(0)$	$m-k$	$L(0)$	$m-k$	$L(0)$
1·25	2·35	1·80	2·57	2·50	2·55	3·50	2·37
1·20	2·50	1·20	3·40	1·50	3·64	3·00	2·48
1·00	2·74	1·00	3·90	1·00	4·90	2·50	2·99
0·80	3·36	0·90	4·17	0·80	5·80	1·50	4·31
0·60	4·15	0·80	4·50	0·60	7·34	1·00	5·90
0·30	5·84	0·60	5·63	0·50	8·70	0·75	7·44
0·20	8·41	0·40	7·40	0·40	10·20	0·50	10·42
0·15	7·54	0·20	10·18	0·30	12·60	0·40	12·13
0·05	9·11	—		—	—	—	—

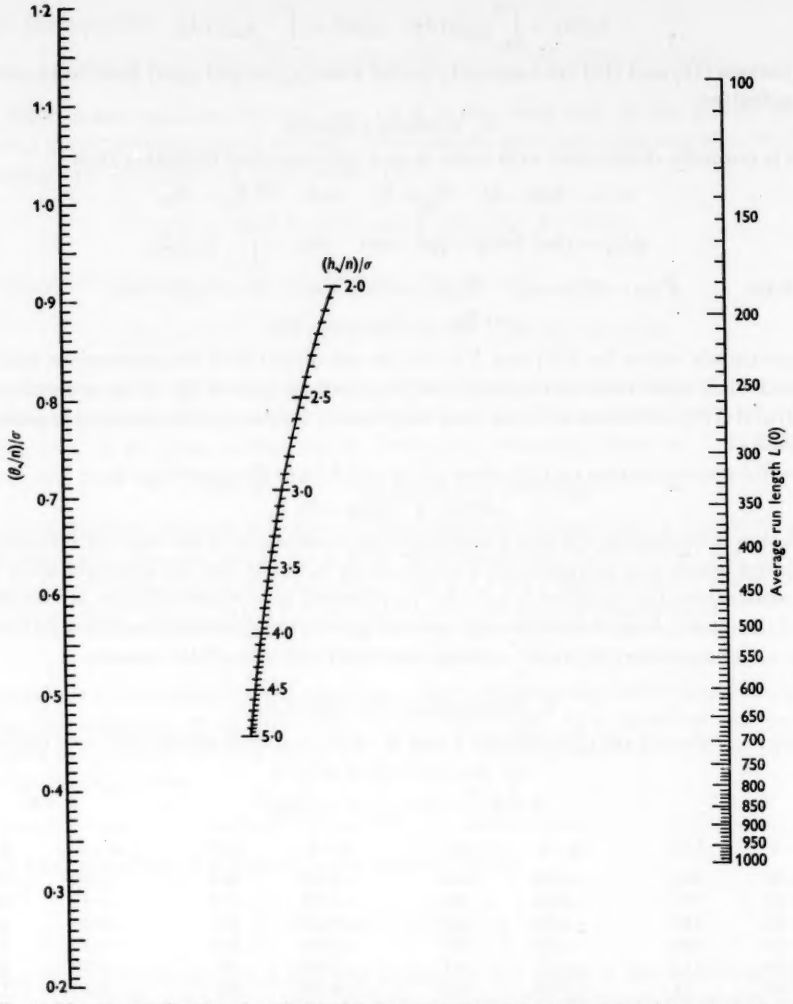


Fig. 4. Nomogram showing the relationship for a normal variate between $(\theta\sqrt{n})/\sigma$, $(h\sqrt{n})/\sigma$ and average run length at A.Q.L. θ = (reference value − A.Q.L.), h = decision interval, n = sample size.

Values of the A.R.L. of a number of schemes for a normal variate with unit variance were obtained by the method just described. The values obtained for $L(0)$ after successive substitution in the appropriate equations are shown in Tables 1 *a* and *b*. From the data given in these tables and the techniques described by Lyle (1954) it is possible to construct nomograms from which the A.R.L.'s of a wide variety of schemes can be obtained at both acceptable and rejectable quality levels (R.Q.L.). A nomogram which has been constructed in this way is shown in Fig. 4. From this diagram the A.R.L.'s of a number of schemes at A.Q.L. can be obtained. A similar one can be drawn for R.Q.L. In Figs. 4 and 5, σ is the standard

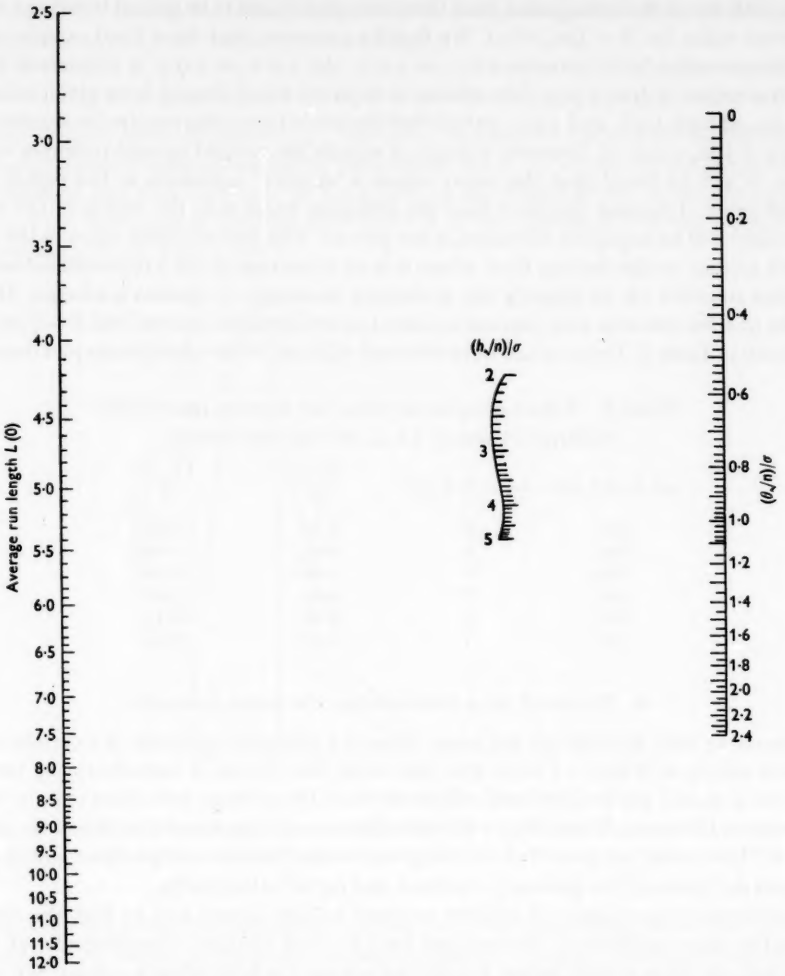


Fig. 5. Nomogram showing the relationship for a normal variate between $(\theta\sqrt{n})/\sigma$, $(h\sqrt{n})/\sigma$ and average run length at R.Q.L. θ = (R.Q.L. − reference value), h = decision interval, n = sample size.

deviation of the population which is being sampled and whose mean value is being controlled, n is the size of the sample which is taken at each test and $\theta = |m-k|$. If m_a is the mean at A.Q.L. the A.R.L. at A.Q.L. is obtained by placing a ruler on the nomogram so that it joins the points $h\sqrt{(n)}/\sigma$ and $\theta\sqrt{(n)}/\sigma = (m_a-k)\sqrt{(n)}/\sigma$, and reading off the run length from the scale on the right-hand side of the diagram. The A.R.L. at R.Q.L. when the mean is m_r is obtained in the same way from Fig. 5.*

A number of schemes can be constructed using different decision intervals and reference values which will have the same A.R.L., either at A.Q.L. or R.Q.L. or at both. It is possible to

* The accuracy obtained from using the nomograms is demonstrated in the Appendix.

show, with use of the nomograms, that there are advantages to be gained by using a central reference value, i.e. $k \doteqdot \frac{1}{2}(m_a + m_r)$. We find for example, that for a fixed sample size, of the schemes which have the same A.R.L. at A.Q.L., the A.R.L. at R.Q.L. is minimized when k is in the region of $\frac{1}{2}(m_a + m_r)$. If a scheme is required which should have given values for the A.R.L. at both A.Q.L. and R.Q.L. we find that the level of sampling required is at a minimum when $k \doteqdot \frac{1}{2}(m_a + m_r)$. If, however, a graph of sample size plotted against reference value is drawn, it will be found that this curve shows a 'shallow' minimum in the region of the central value. Provided therefore that the reference value is in the region of the central value there will be negligible difference in sample size. This has practical value in the actual use of a scheme on the factory floor, where it is an advantage to use a reference value which has been rounded off, to simplify the arithmetic necessary to operate a scheme. The run lengths of a few schemes with various combinations of decision interval and $\theta = \frac{1}{2}|m_a - m_r|$ are shown in Table 2. These values were obtained with use of the nomograms just described.

Table 2. *Values of reference value and decision interval for different choices of* A.R.L. *for a normal variate*

A.R.L. at A.Q.L.	A.R.L. at R.Q.L.	$\frac{\theta\sqrt{n}}{\sigma}$	$\frac{h\sqrt{n}}{\sigma}$
1000	3	1·12	2·40
1000	7	0·65	4·06
500	3	1·04	2·26
500	7	0·60	3·80
250	3	0·94	2·11
250	7	0·54	3·51

8. EXAMPLE OF A SCHEME FOR A NORMAL VARIATE

Suppose we wish to maintain the mean value of a measured property of a certain manufactured article at 5 units or more and that when the process is manufacturing items at this level it should not be interfered with in error on the average more than once in 500 hr. If, however, the mean drops below 4·80 units then on average we wish to detect the change in 6½ hr. Let us also suppose that it is only convenient to test once per hour and that the standard deviation of the process is constant and equal to 0·48 units.

To determine the number of articles we need to test hourly and to find the decision interval we may use Table 2. We find that for A.R.L.'s of 500 and 7, respectively, at A.Q.L. and R.Q.L., $(\theta\sqrt{n})/\sigma = 0{\cdot}60$, where $\theta = \frac{1}{2}$(A.Q.L − R.Q.L.) $= 0{\cdot}10$; thus $n = 8{\cdot}30$. We would therefore take samples of 9; this gives a scheme in which $(\theta\sqrt{n})/\sigma = 0{\cdot}625$. By using the nomograms we find that if the A.R.L. at A.Q.L. is to be 500 then $(h\sqrt{n})/\sigma = 8{\cdot}67$, which gives $h = 0{\cdot}59$ and an A.R.L. at R.Q.L. of 6·8.

9. SAMPLE SIZE AND SAMPLING INTERVAL

From the nomograms of Figs. 4 and 5 it is possible to determine the most suitable values of the reference value and the decision interval to use once the A.Q.L. and R.Q.L. have been defined and the average durations at these qualities have been specified. We find from these nomograms that the relationship between the sampling interval and sample size is very nearly linear so that the difference in average duration at A.Q.L. and R.Q.L between taking one sample every hour, or a sample, three times as big, every third hour is relatively small.

It is therefore best to sample as frequently as practical considerations allow, and to determine the appropriate sample size accordingly.

10. The run-length distribution of a scheme for a normal variate

The run-length distributions at A.Q.L. and R.Q.L. of a scheme for controlling the mean of a normal population were calculated by using equations (17), (18) and (19). The population being sampled had unit variance; at A.Q.L., $m_a - k$ was taken to be $-0{\cdot}50$, and at R.Q.L. $m_r - k = 1{\cdot}00$; the decision interval used was $h = 3$. The shapes of the two distributions obtained are given in Figs. 6 and 7.

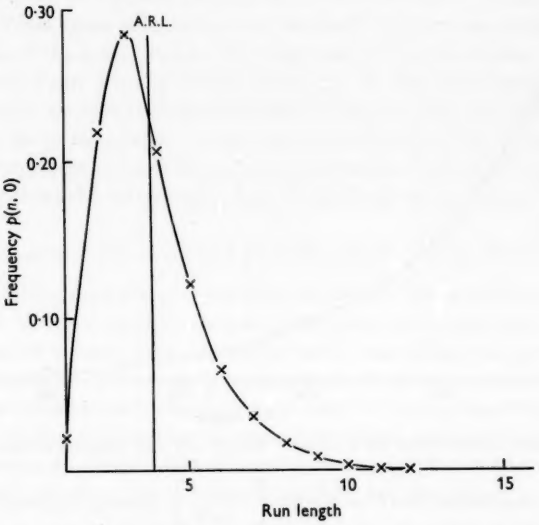


Fig. 6. Run-length distribution for a scheme with $h = 3$ and samples drawn from a normal population with mean = 1·00 and unit variance. Reference value = 0. Average run length = 3·75.

At rejectable quality the distribution has the desirable properties that the probability of obtaining a run length less than the average is high and that the chance of getting a run length greater than three times the average is small. In this particular example the A.R.L. at R.Q.L. is 3·75 and the probability of obtaining a run length of 4 or less is 0·7377. The chance of getting a run length greater than three times the A.R.L. is approximately 1 in 650.

Fig. 7 shows the correspondence between the run-length distribution at A.Q.L. and the negative exponential distribution of equation (15). It is important to recognize that at A.Q.L. the probability of obtaining runs less than the average is high. In our example the chance of getting a run less than 20, i.e. about one-fifth of the average is 1 in 7. The actual value of this probability is 0·1406, whilst the value given by equation (16) is 0·1496. When $L(0)$ is large, the probability of obtaining a run length less than a given multiple of the average value is nearly invariant, since

$$\sum_{1}^{\alpha L(0)} p(n, 0) \doteqdot 1 - e^{-\alpha}. \qquad (21)$$

For high values of $L(0)$ therefore the run-length distribution is specified completely by $L(G)$. It follows that any alternative method of specifying the characteristics of a scheme at A.Q.L. can be expressed in terms of $L(0)$. For example, in some circumstances it may be desirable to ensure that the probability of obtaining a run length at A.Q.L. less than a given figure does not exceed a specified value. The values of $L(0)$ appropriate to this condition can be obtained from equation (16). To take a specific case at A.Q.L., we may wish the probability of obtaining a run length < 50 not to exceed 0·10. This is equivalent to choosing a scheme for which at A.Q.L. $L(0) = 467$. As a further example, we may take the situation where the cost of erroneously concluding that the process is out of control is directly related to the time

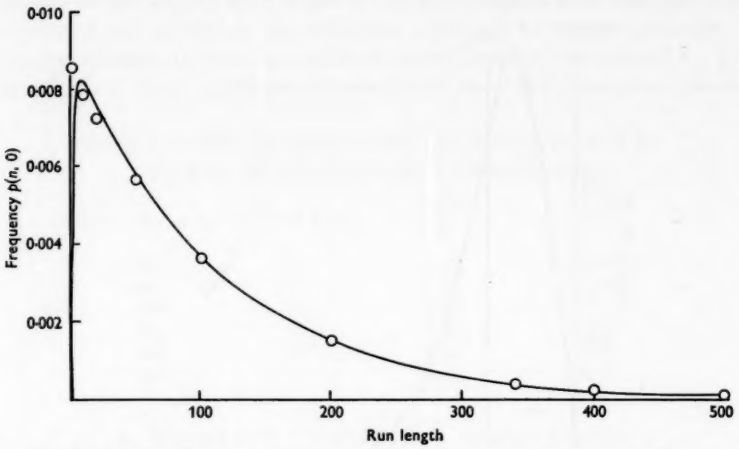


Fig. 7. Run-length distribution for a scheme with $h = 3$ and samples drawn from a normal distribution with mean $= -0{\cdot}50$ and unit variance. Reference value $= 0$. ○ = points on negative exponential distribution $p(n, 0) = \frac{1}{L(0)} \exp\left[-\frac{n-1}{L(0)}\right]$. Average run length = 116·8.

which elapsed since the process was last set in motion. If the cost function $c(n)$ relates the cost of unnecessary process investigation to running time and if the condition is laid down that the average cost due to this type of investigation must not exceed C, then the value of $L(0)$ which will ensure that this is so will be given by the equation

$$C = \sum_{1}^{\infty} \frac{c(n)}{L(0)} \exp\left[-\frac{n-1}{L(0)}\right]. \tag{22}$$

11. CONTROL OF THE STANDARD DEVIATION FOR A NORMAL VARIATE

To control the standard deviation of a process following a normal distribution with use of the range, we can use equations (4) and (5) to calculate the A.R.L.'s of a particular scheme when a function which approximates to the distribution of the sample range is used. For example, we may assume (Cox, 1949) that if x is the sample range or the mean of a number of ranges, $f(x)$ has the form

$$f(x) \doteqdot \frac{(x/a)^{b-1} e^{-x/a}}{a\Gamma(b)},$$

in which case w is the non-zero real root of the equation

$$e^{uk}(1-au)^b = 1. \tag{23}$$

If we use the notation employed in *Tables of the Incomplete Gamma Function* (1957) we have

$$F(\alpha) = I[a/(\alpha\sqrt{b}), b-1], \quad M(\alpha) = abI[\alpha/\{a\sqrt{(b+1)}\}, b-1],$$

$$e^{-wk}G(\alpha) = I\left[\frac{\alpha(1-aw)}{a\sqrt{b}}, b-1\right].$$

These expressions can be used to calculate approximate values of the A.R.L.'s for schemes designed to control the process standard deviation with the use of the range when the number of results available is small and when the above distribution is an adequate approximation to the distribution of the sample range. Davies & Pearson (1934) have given a table of the values of the first two Pearson coefficients of the distribution of the mean of n ranges or r observations. From these calculations we conclude that for the purpose of estimating approximate values of the A.R.L. at A.Q.L. it is reasonable to assume that the *mean* range is normally distributed when $3 \leqslant r \leqslant 10$ and when $n \geqslant 3$. For these values of r and n the nomogram of Fig. 4 can be used to determine schemes for the control of the process standard deviation as well as the process mean. In the figure, therefore, n is the number of ranges used to calculate the mean range and σ is the standard deviation of the distribution of individual ranges which can be obtained, for example, from Table 20 given by Pearson & Hartley (1954).

12. Simultaneous control of the process mean and standard deviation

We have examined the problem of controlling separately the mean or standard deviation of a normal variate. We now consider the case when both these parameters may vary at the same time, but where we are only concerned with controlling the percentage of items outside a specified limit a. If y is a suitable function of the estimated mean and standard deviation we may do this by the method already described by subtracting a reference value k from sample values of y, formulating the cumulative sum and using a decision interval h. A convenient form of y is $y = (a-\bar{x})/s$, where $\bar{x}$ is the sample mean and s is an estimate of the process standard deviation made from the sample. We can write

$$y = \left[\left(\frac{a-\mu}{\sigma}\right) - \left(\frac{\bar{x}-\mu}{\sigma}\right)\right] \Big/ \left(\frac{s}{\sigma}\right),$$

where μ is the process mean and σ its standard deviation. At acceptable quality level $a-\mu = \alpha\sigma$ and at rejectable quality level $a-\mu = \beta\sigma$. If s the estimate of σ is obtained by summing squared deviations about the sample mean, or from the mean of a number of ranges assumed to be distributed as $\chi/\sqrt{\nu}$ based on ν degrees of freedom, y will be distributed as the non-central t. We have not yet been able to apply equations (4) and (5) to obtain approximate values of the A.R.L.'s of different schemes since we have not been able to solve the equation

$$\int_{-\infty}^{+\infty} e^{wy} f(y)\, dy = e^{wk}$$

in a convenient form. We have found for a number of different distributions that if at R.Q.L. the mean value of the parameter being controlled is m_r, the approximation

$$L(0) \doteqdot 1 + h/(m_r - k)$$

for all practical purposes gives a satisfactory estimate of the true value of $L(0)$ over a wide range of values of h and $m_r - k$. The reason for this is fairly obvious. It seems reasonable to

anticipate therefore that in the present instance a close approximation to the A.R.L. at R.Q.L. will be given by the expression $1+h/\theta$, where

$$\theta = [\beta\sqrt{\nu}\,\Gamma(\nu-\tfrac{1}{2})/\{\sqrt{2}\,\Gamma(\tfrac{1}{2}\nu)\}]-k, \tag{24}$$

in which k is the reference value and ν is the effective number of degrees of freedom on which the estimate of the process standard deviation is based. An approximate estimate of the magnitude of the A.R.L. at A.Q.L. can be obtained by finding the A.R.L. of the scheme for a normal variate with equivalent mean and variance. This procedure will often overestimate the actual A.R.L. of the scheme at A.Q.L. since the area in the relevant part of the tail of the normal distribution will usually be less than the corresponding part of the non-central t.

13. POISSON VARIATE

If x is a Poisson variate with mean m, then we can obtain the A.R.L.'s of a scheme by using equations (1) to (5) in which

$$F(\alpha) = \sum_0^\alpha m^x e^{-m}/x!, \quad M(\alpha) = mF(\alpha-1),$$

$$e^{-wk}\,G(\alpha) = \sum_0^\alpha \frac{(m\,e^w)^x}{x!}\exp(-m\,e^w),$$

where w is the real non-zero root of the equation

$$m(e^u-1) = uk. \tag{25}$$

It is convenient at this stage to modify the notation used for $P(z)$, $N(z)$ and $L(z)$. Accordingly $P(z,k,h)$ will be used to denote $P(z)$ for a scheme with reference value k and decision interval h. $N(z,k,h)$ and $L(z,k,h)$ similarly define $N(z)$ and $L(z)$ for the same scheme.

If calculations are undertaken to determine the A.R.L. of a number of schemes with different values of h and k but the same value of m, considerable economies in computing effort can be effected with use of the approximate relationships

$$P(z,k,h) \doteqdot P(z-1,k-1,h+1), \tag{26}$$

$$P(z,k,h) \doteqdot P(z+1,k,h+1). \tag{27}$$

For example, if $m=3$, $z=1$, $k=5$ and $h=5$ we have

$$P(1,5,5) = 0{\cdot}9925, \quad P(0,4,6) = 0{\cdot}9892, \quad P(2,5,6) = 0{\cdot}9918.$$

If $P(z,k,h)$ is known for all z between 0 and h, accurate values of $P(z+1,k,h+1)$ can quickly be obtained by direct substitution in the set of equations generated by equation (1) and using equation (27).

Calculations of the true A.R.L. at A.Q.L. and R.Q.L. of a number of schemes for a Poisson variate were calculated in this way. From these a number of schemes with A.R.L. at A.Q.L. of 500 and A.R.L.'s of 3 and 7 at R.Q.L. were determined by graphical interpolation. We have found these particular values of the A.R.L. to be useful in practice. The values of h, k and m which relate to these schemes are given in Tables 3 and 4. We find from these tables and other calculations that the schemes for which the sample size is smallest are those which have reference values which are in the neighbourhood of the central value. Accordingly, only the schemes with reference values in this region are shown in the tables. It should be

Table 3. *Values of* m_a, m_r, $R = (m_r/m_a)$, h *and* k *for schemes with* A.R.L. = 500 *at* A.Q.L. *and* A.R.L. = 7 *at* R.Q.L. *for a Poisson variate*

R	h	k	m_a	m_r	R	h	k	m_a	m_r
4·64	2	1	0·22	1·02	1·65	9	5	3·84	6·33
3·41	3	1	0·39	1·33	1·64	10	5	3·98	6·52
3·04	4	1	0·52	1·58	1·60	8	6	4·40	7·02
2·90	5	1	0·62	1·80	1·59	9	6	4·60	7·30
2·80	6	1	0·70	1·96	1·58	10	6	4·75	7·50
2·79	7	1	0·77	2·15	1·57	11	6	4·89	7·70
2·43	4	2	1·01	2·45	1·55	9	7	5·33	8·26
2·29	5	2	1·18	2·70	1·54	10	7	5·51	8·46
2·21	6	2	1·31	2·89	1·53	11	7	5·67	8·65
2·15	7	2	1·43	3·07	1·52	9	8	6·07	9·24
2·12	8	2	1·53	3·24	1·50	10	8	6·28	9·45
2·10	9	2	1·62	3·41	1·50	11	8	6·45	9·65
2·07	5	3	1·75	3·62	1·49	12	8	6·58	9·83
1·97	6	3	1·95	3·84	1·48	10	9	7·05	10·44
1·89	7	3	2·13	4·03	1·47	11	9	7·25	10·65
1·87	8	3	2·26	4·22	1·47	12	9	7·38	10·83
1·85	9	3	2·37	4·40	1·46	10	10	7·81	11·40
1·79	7	4	2·79	5·00	1·44	11	10	8·05	11·61
1·76	8	4	2·96	5·20	1·44	12	10	8·20	11·81
1·74	9	4	3·10	5·39	1·42	11	11	8·85	12·61
1·72	10	4	3·21	5·52	1·42	12	11	9·00	12·81
1·67	8	5	3·68	6·16					

Table 4. *Values of* m_a, m_r, $R = (m_r/m_a)$, h *and* k *for schemes with* A.R.L. = 500 *at* A.Q.L. *and* A.R.L. = 3 *at* R.Q.L. *for a Poisson variate*

R	h	k	m_a	m_r	R	h	k	m_a	m_r
7·77	2	1	0·22	1·71	2·26	5	6	3·57	8·06
5·64	3	1	0·39	2·20	2·18	6	6	3·90	8·50
5·19	4	1	0·52	2·70	2·14	7	6	4·17	8·91
4·98	5	1	0·62	3·09	2·12	8	6	4·40	9·31
4·05	3	2	0·79	3·20	2·07	6	7	4·60	9·50
3·63	4	2	1·01	3·67	2·04	7	7	4·87	9·92
3·43	5	2	1·18	4·05	2·02	8	7	5·11	10·33
3·34	3	3	1·23	4·11	2·00	6	8	5·24	10·50
2·97	4	3	1·52	4·51	1·96	7	8	5·57	10·91
2·87	5	3	1·75	5·03	1·95	8	8	5·83	11·38
2·83	6	3	1·95	5·51	1·94	9	8	6·07	11·80
2·77	7	3	2·13	5·91	1·91	7	9	6·25	11·92
2·75	4	4	2·04	5·60	1·89	8	9	6·55	12·39
2·58	5	4	2·33	6·02	1·87	9	9	6·83	12·79
2·52	6	4	2·58	6·50	1·86	7	10	6·94	12·93
2·48	7	4	2·79	6·91	1·84	8	10	7·28	13·39
2·47	8	4	2·96	7·31	1·82	9	10	7·57	13·80
2·43	5	5	2·92	7·09	1·80	8	11	8·00	14·40
2·35	6	5	3·21	7·53	1·78	9	11	8·30	14·80
2·30	7	5	3·46	7·96	1·77	10	11	8·59	15·23
2·27	8	5	3·68	8·34					

noted that it is not possible to interpolate in these tables for non-integral values of h and k.

Situations arise in which the value of the mean is small and the number of items which it is possible to test is also small, so that the values which m_a and m_r take are such that schemes

with fractional values of h and k have to be used. If $L(0, k, h)$ is the A.R.L. of a scheme in which h and k are integers and if Δh represents a small non-integral increment in h, then

$$L(0, k, h+\Delta h) = L(0, k, h+1).$$

A discontinuity occurs in $L(0, k, h)$ at h.

For schemes where it is necessary to use fractional values of h, the equations satisfied by $P(z, k, h)$ and $N(z, k, h)$ must be written in forms which take into account the discontinuous nature of the distribution function. For example, we write

$$P(z, k, h) = \sum_{0}^{[k-z]} \frac{m^w e^{-m}}{w!} + \sum_{[k-z]+1}^{\{k+h-z\}} P(w+z-h, k, h) \frac{m^w e^{-m}}{w!}, \tag{28}$$

where $[x]$ denotes the nearest integer which is $\leqslant x$ and $\{x\}$ denotes the nearest integer y which is such that $y < x$. A similar equation holds for $N(z, k, h)$. $P(z, k, h)$, $N(z, k, h)$ and $L(z, k, h)$ are discontinuous functions of z, k and h. We have already remarked upon the discontinuity in $L(0, k, h)$ when h is increased by a small amount from an integral value. It is not difficult to show that discontinuities in $L(0, k, h)$ also occur at integral values of k. It can be shown that when h is an integer

$$P(0, k+\Delta k, h) \doteqdot P(0, k, h+1), \quad P(0, k-\Delta k, h) \doteqdot P(0, k, h).$$

Examples of the order of the changes in A.R.L. which occur due to these considerations are shown in Table 5 for the case where $m = 0{\cdot}10$ and $h = 2$.

Table 5. *Values of $L(0, k, h)$ for different k when $m = 0{\cdot}10$ and $h = 2$*

k	$L(0, k, h)$	k	$L(0, k, h)$
0·20	146	0·31	365
0·30	350	0·32	366
0·40	697	0·33	366
0·45	886	0·34	654
0·50	888	0·35	654
0·55	2958	0·36	657
0·60	3022	0·37	666
0·70	5082	0·38	694
1·00	5646	0·40	697

Values of the mean for a Poisson variate, which give an A.R.L. of 500 at A.Q.L. and 7 at R.Q.L. have been obtained for a number of schemes which use fractional values of the reference value k and the decision interval h. These have been calculated for situations in which the sample size for some reason or other must be small. The values of m_a and m_r have been obtained as above by graphical interpolation after run lengths of various values of m had been computed. Values of the A.R.L. at A.Q.L. were obtained by the tedious method of enumerating all the possible paths which cross either decision boundary. The A.R.L. at R.Q.L. was obtained by substituting the approximation

$$L(z, h, k) \doteqdot 1 + (h-z)/(m_r - k) \tag{29}$$

into the equation for $L(0, h, k)$. The appropriate values of h, k used on the right-hand side of the equation are determined with due regard to the discontinuous nature of $L(z, h, k)$. In other words the values of h and k used to determine $L(z, h, k)$ would not in general be the same on both sides of the equation. To take a simple example which makes this more

explicit, if we wished to estimate $L(z, k, 3{\cdot}10)$ when z and k are integers, the value of h which we would take for $1+(h-z)/(m-k)$ would be 4 and not 3·10.

The results obtained from these calculations are given in Table 6.

Table 6. *Values of m_a, m_r, $R = (m_r/m_a)$, h and k for schemes with non-integral values of h and k which have* A.R.L. *at* A.Q.L. $= 500$ *and* A.R.L. *at* R.Q.L. $= 7$

R	h	k	m_a	m_r	R	h	k	m_a	m_r
5·11	4·00	0·30	0·19	0·97	3·70	3·75	0·60	0·33	1·22
5·07	3·00	0·30	0·15	0·76	3·59	3·00	0·70	0·32	1·15
5·06	3·25	0·30	0·16	0·81	3·54	3·50	0·70	0·35	1·24
5·06	3·50	0·30	0·17	0·86	3·54	3·00	0·80	0·35	1·24
5·00	3·75	0·30	0·18	0·90	3·53	3·25	0·70	0·34	1·20
4·78	3·00	0·40	0·18	0·86	3·50	4·00	0·60	0·34	1·22
4·57	3·00	0·50	0·21	0·96	3·47	3·25	0·80	0·38	1·32
4·40	3·50	0·40	0·22	0·97	3·42	3·75	0·70	0·38	1·30
4·38	3·25	0·40	0·21	0·92	3·41	3·00	0·90	0·39	1·33
4·38	4·00	0·40	0·24	1·05	3·33	4·00	0·70	0·40	1·33
4·34	3·75	0·40	0·23	1·00	3·33	3·25	0·90	0·42	1·40
4·20	3·25	0·50	0·25	1·05	3·31	3·50	0·80	0·41	1·36
4·20	3·50	0·50	0·25	1·05	3·23	3·75	0·80	0·43	1·39
4·11	3·75	0·50	0·28	1·15	3·23	3·75	0·90	0·49	1·49
4·11	4·00	0·50	0·28	1·15	3·20	4·00	0·80	0·45	1·44
3·81	3·00	0·60	0·27	1·03	3·10	3·50	0·90	0·46	1·43
3·74	3·50	0·60	0·31	1·16	3·00	4·00	0·90	0·51	1·53
3·73	3·25	0·60	0·30	1·12					

14. Example of a scheme for a Poisson variate

Suppose we wish to control the level of faults which can be seen in lengths of fabric and that quality is acceptable if the faults do not exceed a mean of 5 per 10^6 yards and rejectable if the mean number of faults per 10^6 yards exceeds 10. We suppose also that we require the A.R.L. at A.Q.L to be 500 and the A.R.L. at R.Q.L to be 7. We can find the decision interval, the reference value and the number of yards of fabric we need to examine at a time by using Table 3. We find the values of k and h which give us a ratio $R =$ (R.Q.L./A.Q.L.). If, in the table, we find more than one scheme with this value of R, we choose that one for which m_a is a minimum. For this example, the nearest value of R is 1·97 with $h = 6$ and $k = 3$, $m_a = 1{\cdot}95$ and $m_r = 3{\cdot}84$. For this scheme the number of yards of fabric we need to examine is $(1{\cdot}95/5) \times 10^6$ yards $= 0{\cdot}39 \times 10^6$ yards.

15. Fraction defective inspection schemes

Tables 3, 4 and 6 can also be used to formulate schemes for the control of the proportion of defective items being produced, provided this proportion is very small so that the Poisson distribution may be used to describe the sampling distribution of the number of defective items which will occur in an r-fold sample.

16. Example of a scheme for controlling the properties of defective items being produced

We may take the example where p_a, the proportion of defective items which can be tolerated at A.Q.L. is 0·0050, where p_r the proportion of defective items at R.Q.L. is 0·0150, where the A.R.L. at A.Q.L. is to be 500 and that at R.Q.L. to be 7, and where $rp_a = m_a$ and

$rp_r = m_r$, so that $m_r/m_a = p_r/p_a = 3{\cdot}00$. The nearest value of R we obtain to this from Table 3 is $R = 3{\cdot}04$ when $k = 1$ and $h = 4$, $m_a = 0{\cdot}52$ so that $r = 104$. We would therefore take samples of 104 at regular intervals, count the number of defectives, form the cumulative sum about a reference value of 1 and take action if this becomes $\geqslant 4$.

The authors wish to thank the Directors of British Nylon Spinners Limited for permission to publish this paper. They are also grateful to Professor G. A. Barnard for useful discussions and to Dr E. S. Page for suggesting several improvements to the original draft.

REFERENCES

BARNARD, G. A. (1954). Sampling inspection and statistical decisions. *J. R. Statist. Soc.* B, **16**, 151–65.

BARNARD, G. A. (1959). Control charts and stochastic processes. *J. R. Statist. Soc.* B, **21**, 239–57.

BARRACLOUGH, E. D. & PAGE, E. S. (1959). Tables for Wald tests for the mean of a normal distribution. *Biometrika*, **46**, 169–77.

COX, D. R. (1949). The use of range in sequential analysis. *J. R. Statist. Soc.* B, **11**, 101–14.

DAVIES, O. L. & PEARSON, E. S. (1934). Methods of estimating from samples the population standard deviation. *J. R. Statist. Soc.* (Suppl.), **1**, 76–93.

GIRSHICK, M. A. (1946). Continuations to the theory of sequential analysis, II and III. *Ann. Math. Statist.* **17**, 282–98.

KEMP, K. W. (1958). Formulae for calculating the operating characteristics and the average sample number of some sequential tests. *J. R. Statist. Soc.* B, **20**, 379–86.

LYLE, P. (1954). The construction of nomograms for use in statistics. Part I. *Appl. Statist.* **3**, 116–25.

PAGE, E. S. (1954). Continuous inspection schemes. *Biometrika*, **41**, 100–15.

PAGE, E. S. (1957). On problems in which a change in a parameter occurs at an unknown point. *Biometrika*, **44**, 248–52, 248.

PEARSON, E. S. & HARTLEY, H. O. (1954). *Biometrika Tables for Statisticians*, **1**. Cambridge University Press.

Tables of the Incomplete Gamma Function (1957). Edited by Karl Pearson. Cambridge University Press.

APPENDIX (*added in proof*)

Barraclough & Page (1959) have published exact computations of the values of the various parameters for a number of Wald sequential tests. From these it is possible in a few instances to compare values of $L(0)$ obtained from the nomograms shown in Figs. 4 and 5 with exact values obtained with use of the tables given at the end of their paper. The table below shows the comparisons it was possible to make, the columns headed B. & P. show the vales of $L(0)$ obtained from the work of these authors whilst the values obtained from the nomograms are shown in the columns headed 'Nom.'.

Comparison between values of $L(0)$ obtained from Barraclough & Page's calculations and those obtained from the nomogram

		$L(0)$			$L(0)$	
h	$m-k$	B. & P.	Nom.	$m-k$	B. & P.	Nom.
4·10	−0·50	376	370	0·50	8·92	8·85
2·70	−0·75	282	274	0·75	4·28	4·32
3·21	−0·75	580	580	0·75	4·99	5·05
3·50	−0·75	912	920	0·75	5·37	5·42
1·97	−1·00	246	240	1·00	2·71	2·70

Biometrika (1960), **47**, 3 *and* 4, *p.* 381
With 1 *text-figure*
Printed in Great Britain

Table of Neyman-shortest unbiased confidence intervals for the binomial parameter†

BY COLIN R. BLYTH AND DAVID W. HUTCHINSON
University of Illinois, Urbana, Illinois

1. *Introduction and summary.* A random subset A of the possible values of a parameter θ is by definition an unbiased α-confidence set for θ if

$$P_\theta(\theta \in A) \geqslant \alpha, \quad \text{all } \theta$$
$$P_\theta(\theta' \in A) \leqslant P_\theta(\theta \in A), \quad \text{all pairs } \theta, \theta'.$$

A uniformly Neyman-shortest member of this class is one which minimizes $P_\theta(\theta' \in A)$ for all pairs θ, θ'. Such a set is optimum in the sense that among all unbiased α-confidence sets it uniformly minimizes the probability of covering false values. These concepts were introduced and construction methods given by Neyman (1937). When such a confidence set is to be based on a discrete random variable, randomization is needed for its construction; see Eudey (1949), Pearson (1950), Tocher (1950) and Stevens (1950). The Neyman-shortest unbiased confidence set for the binomial parameter was obtained by Eudey (1949); the purpose of the present note is to provide tables of this set.

2. *Use of the tables.* Let X be binomial (n, p) and Y be rectangular $(0, 1)$ with X and Y independent. If $X = x$ and $Y = y$ (using a table of random numbers) are observed, enter the α-table at the location $n, x+y$ and read there the observed α-confidence set for p. If the n, $x+y$ location is not tabled, enter the table at location n, $n+1-(x+y)$ and use the reflexion about $p = \frac{1}{2}$ of the set given there. Tables are given for $\alpha = 0{\cdot}95, 0{\cdot}99$ and $n = 2(1)\,24(2)\,50$. Intervals for $n = 25, 27, \ldots, 49$ can be found by interpolation. The sets are intervals whose end-points are given to two decimal places; rounding-off errors are small compared to the lengths of the intervals.

Example. If $n = 21$, $X = 12$, $Y = 0{\cdot}73$, enter the table at $n = 21$, $n+1-(X+Y) = 9{\cdot}27$. Interpolating, we see that the 0·99 confidence interval is $1-0{\cdot}69$, $1-0{\cdot}17 = 0{\cdot}31, 0{\cdot}83$.

3. *Comparison with existing tables* The confidence intervals of Clopper & Pearson (1934) are obtained from tests whose rejection regions consist of two tails with approximate probability $\frac{1}{2}(1-\alpha)$ each. These confidence sets are non-randomized and so give coverage probabilities in excess of α at the expense of being larger than necessary to achieve probability α of coverage. The corresponding randomized intervals based on tests whose rejection regions consist of two tails with exact probability $\frac{1}{2}(1-\alpha)$ each are easily computed—see Stevens (1950)—using a desk calculator and tables of the binomial distribution. These confidence sets are very reasonable and though they have no particular optimum property, are not likely to be much improved upon.

The confidence sets given by Crow (1956) are based on tests whose acceptance regions are made up of the most probable points. The resulting confidence sets have the property that among all non-randomized confidence intervals they minimize the sum of the $n+1$ possible lengths.

† Work supported by the National Science Foundation.

The Eudey intervals tabled here include the advantage of randomization and are optimum in the sense of being Neyman-shortest unbiased.

In the following figure the Clopper-Pearson, Stevens, Crow and Eudey 0·95-confidence intervals are compared for $n = 10$. Corresponding to a given $X + Y$ value the various confidence limits are read off as vertical intercepts.

This comparison shows the Stevens interval to be shorter than Eudey's for

$$1{\cdot}6 \leqslant X + Y \leqslant 9{\cdot}4$$

and the reverse for other values of $X + Y$. Since shortness in the sense of ordinary length is not necessarily desirable (it seems undesirable when the interval fails to cover p) corresponding preferences for Stevens v. Eudey intervals do not necessarily follow.

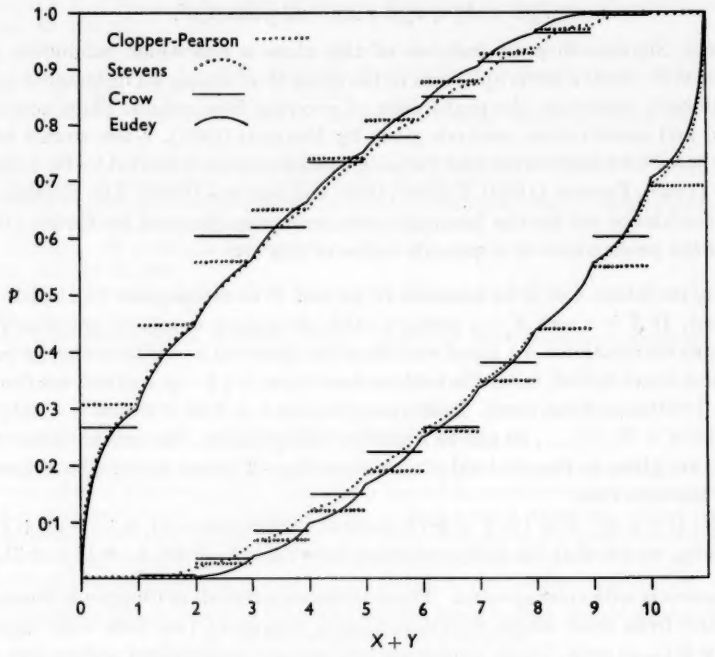


Fig. 1. 0·95-confidence intervals for p based on X, binomial $(10, p)$ and an independent Y, rectangular $(0, 1)$.

4. *Construction of the tables* Let X be Binomial (n, p) and Y be Rectangular $(0, 1)$, with X and Y independent. Eudey (1949) showed that a uniformly most powerful unbiased level $1-\alpha$ test of $p = p^* \, vs. \, p \neq p^*$, based on X, is given by the acceptance region $a(p^*)$ determined by

$$a(p^*) = \{x, y : n_0 + \gamma_0 \leqslant X + Y \leqslant n_1 + \gamma_1\},$$

$$n_0, n_1 \text{integers}, \quad 0 \leqslant \gamma_0 \leqslant 1, \quad 0 \leqslant \gamma_1 \leqslant 1,$$

$$[P_p(n_0 + \gamma_0 \leqslant X + Y \leqslant n_1 + \gamma_1)]_{p=p^*} = \alpha,$$

$$[(d/dp) P_p(n_0 + \gamma_0 \leqslant X + Y \leqslant n_1 + \gamma_1)]_{p=p^*} = 0.$$

These equations for n_0, n_1, γ_0, γ_1 reduce to:

$$n_0, n_1 \text{ integers}; \quad 0 \leqslant \gamma_0 \leqslant 1, \quad 0 \leqslant \gamma_1 \leqslant 1,$$

$$\gamma_0 = \frac{(n_1 - np^*)\{P_{p^*}(n_0 \leqslant X \leqslant n_1 - 1) - \alpha\} - n_0(1-p^*)\,P_{p^*}(X = n_0) + n_1(1-p^*)\,P_{p^*}(X = n_1)}{(n_1 - n_0)\,P_{p^*}(X = n_0)},$$

$$\gamma_1 = \frac{(n_0 - np^*)\{P_{p^*}(n_0 \leqslant X \leqslant n_1 - 1) - \alpha\} - n_0(1-p^*)\,P_{p^*}(X = n_0) + n_1(1-p^*)\,P_{p^*}(X = n_1)}{(n_1 - n_0)\,P_{p^*}(X = n_1)}.$$

The unique solvability of these equations follows from obvious continuity and monotonicity considerations. The Neyman–Pearson lemma with side conditions shows that the uniformly most powerful, level $1-\alpha$, test is unique except for the way in which randomization is carried out. The equations are solved by using trial values for n_0, n_1 until the resulting γ_0, γ_1 are both on $[0, 1]$. Solution of the equations was carried out using the University of Illinois Digital Computer Laboratory's ILLIAC. A program was constructed which for arbitrary n, α prints out $n_0 + \gamma_0$, $n_1 + \gamma_1$ for any equally spaced set of p^* values. The p^* values used were 0·005, 0·015, ..., 0·995. The Neyman-shortest unbiased α-confidence set

$$A(X, Y) = \{p^*: X + Y \in a(p^*)\}$$

for p, can then be read off to two places from the $n_0 + \gamma_0$, $n_1 + \gamma_1$ print-out.

Tables of Neyman-shortest unbiased α-confidence sets for a Poisson parameter are being constructed in the same way.

References

Clopper, C. J. & Pearson, E. S. (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*, **26**, 404.

Crow, E. L. (1956). Confidence intervals for a proportion. *Biometrika*, **43**, 423.

Eudey, M. W. (1949). On the treatment of discontinuous random variables. *Technical Report No.* **13**, Statistical Laboratory, University of California, Berkeley.

Neyman, J. (1937). Outline of a theory of statistical estimation based on the classical theory of probability. *Phil. Trans.* A, **236**, 333.

Pearson, E. S. (1950). On questions raised by the combination of tests based on discontinuous distributions. *Biometrika*, **37**, 383.

Stevens, W. L. (1950). Fiducial limits of the parameter of a discontinuous random variable. *Biometrika*, **37**, 117.

Tocher, K. D. (1950). Extension of Neyman–Pearson theory of testing hypotheses to discontinuous variates. *Biometrika*, **37**, 130.

Table 1. *Neyman-shortest unbiased 0·95-confidence intervals for p based on X, Binomial (n, p) and an independent Y, Rectangular (0, 1)*

$n=2$		$n=3$		$n=4$		$n=5$		$n=6$		$n=7$		$n=8$		$n=9$		$n=10$		$X+Y$
00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	0
00	50	00	29	00	23	00	18	00	15	00	13	00	11	00	10	00	09	0·1
00	75	00	50	00	37	00	31	00	26	00	23	00	20	00	18	00	16	·2
00	83	00	59	00	45	00	38	00	32	00	28	00	25	00	22	00	20	·3
00	87	00	65	00	50	00	42	00	35	00	31	00	28	00	25	00	23	·4
00	90	00	68	00	54	00	44	00	38	00	34	00	30	00	27	00	25	·5
00	92	00	71	00	56	00	46	00	41	00	36	00	32	00	29	00	26	·6
00	93	00	73	00	59	00	48	00	42	00	37	00	33	00	30	00	27	·7
00	94	00	75	00	60	00	50	00	44	00	39	00	34	00	31	00	29	·8
00	94	00	76	00	62	00	51	00	45	00	40	00	35	00	32	00	30	·9
00	95	00	78	00	63	00	53	00	46	00	41	00	36	00	33	00	30	1·0
00	1·0	00	85	00	69	00	58	00	50	00	44	00	40	00	36	00	33	1·1
00	1·0	00	89	00	73	00	62	00	53	00	47	00	42	00	38	00	35	1·2
00	1·0	00	92	00	77	00	65	00	56	00	50	00	45	00	40	00	37	1·3
00	1·0	00	94	00	79	00	67	00	58	00	52	00	46	00	42	00	39	1·4
00	1·0	00	95	00	81	00	69	00	60	00	53	00	48	00	44	00	40	1·5
		00	96	00	83	00	71	00	61	00	55	00	49	00	45	00	41	1·6
		00	96	00	84	00	72	00	63	00	56	00	50	00	46	00	42	1·7
		00	97	00	85	00	73	00	64	00	57	00	51	00	47	00	43	1·8
		00	97	00	86	00	74	00	65	00	58	00	52	00	48	00	44	1·9
		03	97	02	86	01	75	01	66	01	58	01	53	01	48	01	44	2·0
				02	90	01	79	01	69	01	61	01	55	01	50	01	46	2·1
				02	93	02	81	01	71	01	63	01	57	01	52	01	48	2·2
				02	95	02	83	01	73	01	65	01	59	01	54	01	50	2·3
				03	96	02	85	02	75	01	67	01	60	01	55	01	51	2·4
				03	97	03	86	02	77	02	68	01	62	01	57	01	52	2·5
						03	88	02	78	02	69	02	63	02	58	01	53	2·6
						04	88	03	79	03	70	02	63	02	59	02	54	2·7
						05	89	04	80	03	71	03	64	03	59	02	55	2·8
						07	90	06	80	05	72	04	65	03	60	03	55	2·9
						10	90	08	81	06	73	05	66	05	61	04	56	3·0
								08	84	07	75	06	68	05	62	04	58	3·1
								09	86	07	77	06	70	05	64	05	59	3·2
								09	87	07	79	06	71	06	65	05	60	3·3
								10	88	08	80	07	73	06	66	05	62	3·4
								11	89	09	81	07	74	06	67	06	63	3·5
										09	82	08	75	07	68	06	63	3·6
										11	83	09	75	08	69	07	64	3·7
										12	84	10	76	09	70	08	65	3·8
										13	84	11	77	10	70	09	65	3·9
										15	85	13	77	11	71	10	66	4·0
												13	79	12	73	10	67	4·1
												14	81	12	74	11	69	4·2
												14	82	12	75	11	70	4·3
												15	83	13	77	11	71	4·4
												16	84	14	77	12	72	4·5
														14	78	13	72	4·6
														15	79	14	73	4·7
														16	80	14	74	4·8
														18	80	16	74	4·9
														19	81	17	75	5·0
																17	76	5·1
																18	78	5·2
																18	79	5·3
																19	79	5·4
																20	80	5·5

Note: the pairs of figures are lower and upper confidence limits for p, given to two places of decimals.

Table 1 (*cont.*)

$n = 11$		$n = 12$		$n = 13$		$n = 14$		$n = 15$		$n = 16$		$n = 17$		$n = 18$		$n = 19$		$X+Y$
00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	01	00	0	0
00	08	00	08	00	07	00	07	00	06	00	06	00	05	00	05	00	05	0·1
00	15	00	14	00	13	00	12	00	11	00	10	00	10	00	09	00	09	·2
00	19	00	17	00	16	00	15	00	14	00	13	00	12	00	12	00	11	·3
00	21	00	19	00	18	00	17	00	16	00	15	00	14	00	13	00	13	·4
00	23	00	21	00	19	00	18	00	17	00	16	00	15	00	14	00	14	·5
00	24	00	22	00	21	00	19	00	18	00	17	00	16	00	15	00	15	·6
00	25	00	23	00	22	00	20	00	19	00	18	00	17	00	16	00	15	·7
00	26	00	24	00	23	00	21	00	20	00	19	00	18	00	17	00	16	·8
00	27	00	25	00	23	00	22	00	21	00	19	00	18	00	17	00	17	·9
00	28	00	26	00	24	00	22	00	21	00	20	00	19	00	18	00	17	1·0
00	32	00	30	00	28	00	26	00	25	00	23	00	22	00	21	00	20	1·2
00	36	00	33	00	31	00	29	00	27	00	26	00	24	00	23	00	22	1·4
00	38	00	35	00	33	00	31	00	29	00	27	00	26	00	25	00	23	1·6
00	40	00	37	00	34	00	32	00	30	00	29	00	27	00	26	00	25	1·8
01	41	00	38	00	36	00	34	00	32	00	30	00	28	00	27	00	26	2·0
01	45	01	41	01	39	00	36	00	34	00	32	00	31	00	29	00	28	2·2
01	47	01	44	01	41	01	39	01	36	01	34	01	32	00	31	00	29	2·4
01	49	01	46	01	43	01	40	01	38	01	36	01	34	01	32	01	31	2·6
02	51	02	47	02	44	02	42	01	39	01	37	01	35	01	34	01	32	2·8
04	52	03	49	03	45	03	43	03	40	02	38	02	36	02	34	02	33	3·0
04	55	04	51	03	48	03	45	03	43	03	40	03	38	02	36	02	35	3·2
05	57	04	53	04	50	04	47	03	44	03	42	03	40	03	38	03	36	3·4
06	59	05	55	05	52	04	49	04	46	04	44	03	41	03	39	03	38	3·6
07	60	06	56	06	53	05	50	05	47	04	45	04	42	04	40	04	39	3·8
09	61	08	58	07	54	07	51	06	48	06	46	05	43	05	41	05	39	4·0
09	64	08	60	08	56	07	53	07	50	06	48	06	45	05	43	05	41	4·2
10	66	09	62	08	58	08	55	07	52	07	49	06	47	06	45	06	43	4·4
11	68	10	63	09	60	09	56	08	53	07	51	07	48	06	46	06	44	4·6
13	69	12	65	11	61	10	57	09	54	08	52	08	49	07	47	07	45	4·8
15	70	14	66	12	62	11	58	10	55	10	53	09	50	08	48	08	46	5·0
16	72	14	68	13	64	12	60	11	57	10	54	10	52	09	49	08	47	5·2
17	74	15	70	14	66	13	62	12	59	11	56	10	53	10	51	09	48	5·4
18	76	16	71	15	67	14	63	13	60	12	57	11	54	10	52	10	50	5·6
20	77	18	72	16	68	15	64	14	61	13	58	12	55	11	53	11	50	5·8
22	78	20	73	18	69	17	65	15	62	14	59	13	56	12	54	12	51	6·0
		21	75	19	71	17	67	16	64	15	61	14	58	13	55	12	53	6·2
		22	77	20	73	18	69	17	65	16	62	14	59	14	56	13	54	6·4
		23	78	21	74	19	70	18	66	16	63	15	60	14	58	14	55	6·6
				23	75	21	71	19	67	18	64	16	61	15	58	14	56	6·8
				24	76	22	72	21	68	19	65	18	62	17	59	16	57	7·0
						23	74	21	70	20	67	18	63	17	61	16	58	7·2
						24	75	22	71	21	68	19	65	18	62	17	59	7·4
						25	76	23	72	21	69	20	66	19	63	18	60	7·6
								24	73	23	70	21	67	20	64	18	61	7·8
								26	74	24	71	22	68	21	65	20	62	8·0
										25	72	23	69	21	66	20	63	8·2
										26	73	24	70	22	67	21	64	8·4
										27	74	25	71	23	68	22	65	8·6
												26	72	24	69	23	66	8·8
												27	73	26	70	24	67	9·0
														26	71	25	68	9·2
														27	72	25	69	9·4
														28	73	26	70	9·6
																27	71	9·8
																28	72	10·0

Note: the pairs of figures are lower and upper confidence limits for p, given to two places of decimals.

Table 1 (*cont.*). 0·95 *confidence intervals*

$n = 20$		$n = 21$		$n = 22$		$n = 23$		$n = 24$		$n = 26$		$n = 28$		$n = 30$		$n = 32$		$X+Y$
00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	0
00	05	00	04	00	04	00	04	00	04	00	04	00	03	00	03	00	03	0·1
00	08	00	08	00	08	00	07	00	07	00	07	00	06	00	06	00	05	·2
00	11	00	10	00	10	00	09	00	09	00	08	00	08	00	07	00	07	·3
00	12	00	11	00	11	00	10	00	10	00	09	00	09	00	08	00	08	·4
00	13	00	12	00	12	00	11	00	11	00	10	00	09	00	09	00	08	·5
00	14	00	13	00	13	00	12	00	12	00	11	00	10	00	09	00	09	·6
00	15	00	14	00	13	00	13	00	12	00	11	00	11	00	10	00	09	·7
00	15	00	15	00	14	00	13	00	13	00	12	00	11	00	10	00	10	·8
00	16	00	15	00	14	00	14	00	13	00	12	00	12	00	11	00	10	·9
00	16	00	16	00	15	00	14	00	14	00	13	00	12	00	11	00	10	1·0
00	19	00	18	00	17	00	17	00	16	00	15	00	14	00	13	00	12	1·2
00	21	00	20	00	19	00	18	00	18	00	16	00	15	00	14	00	13	1·4
00	22	00	21	00	20	00	20	00	19	00	18	00	16	00	15	00	14	1·6
00	24	00	23	00	22	00	21	00	20	00	19	00	17	00	16	00	15	1·8
00	24	00	23	00	22	00	22	00	21	00	19	00	18	00	17	00	16	2·0
00	27	00	25	00	24	00	23	00	22	00	21	00	20	00	18	00	17	2·2
00	28	00	27	00	26	00	25	00	24	00	22	00	21	00	19	00	18	2·4
01	29	01	28	01	27	01	26	01	25	00	23	00	22	00	20	00	19	2·6
01	31	01	29	01	28	01	27	01	26	01	24	01	23	01	21	01	20	2·8
02	31	02	30	02	29	02	28	02	27	01	25	01	23	01	22	01	20	3·0
02	33	02	32	02	31	02	29	02	28	02	26	01	25	01	23	01	22	3·2
02	35	02	33	02	32	02	31	02	29	02	27	02	26	02	24	01	23	3·4
03	36	03	34	03	33	02	32	02	31	02	28	02	27	02	25	02	23	3·6
03	37	03	35	03	34	03	33	03	31	03	29	02	27	02	26	02	24	3·8
04	38	04	36	04	35	04	33	04	32	03	30	03	28	03	26	03	25	4·0
05	39	05	38	04	36	04	35	04	34	04	31	03	29	03	27	03	26	4·2
05	41	05	39	05	37	04	36	04	35	04	32	04	30	03	28	03	27	4·4
06	42	05	40	05	39	05	37	05	36	04	33	04	31	04	29	03	28	4·6
07	43	06	41	06	39	06	38	05	36	05	34	05	32	04	30	04	28	4·8
08	44	07	42	07	40	06	39	06	37	06	35	05	32	05	31	05	29	5·0
08	45	08	43	07	42	07	40	07	39	06	36	06	34	05	32	05	30	5·2
08	46	08	45	08	43	07	41	07	40	06	37	06	35	05	33	05	31	5·4
09	47	09	46	08	44	08	42	07	41	07	38	06	35	06	33	05	31	5·6
10	48	09	46	09	45	09	43	08	41	07	39	07	36	06	34	06	32	5·8
11	49	10	47	10	45	09	44	09	42	08	39	08	37	07	35	07	33	6·0
12	52	12	50	11	48	11	46	10	45	09	42	09	39	08	37	07	35	6·5
15	54	14	52	13	50	13	48	12	47	11	44	10	41	09	38	09	36	7·0
16	57	15	55	15	53	14	51	13	49	12	46	11	43	10	41	10	38	7·5
18	59	17	57	17	55	16	53	15	51	14	48	13	45	12	42	11	40	8·0
20	62	19	60	18	58	17	56	16	54	15	50	14	47	13	44	12	42	8·5
23	64	21	62	20	60	19	57	18	55	17	52	15	49	14	46	13	43	9·0
24	67	23	64	22	62	21	60	20	58	18	54	17	51	15	48	14	45	9·5
27	69	25	66	24	64	23	62	22	60	20	56	18	52	17	49	16	47	10·0
28	72	27	69	25	66	24	64	23	62	21	58	19	55	18	51	17	49	10·5
		29	71	28	68	26	66	25	64	23	60	21	56	20	53	18	50	11·0
				29	71	28	68	27	66	24	62	22	58	21	55	19	52	11·5
						30	70	29	67	26	63	24	60	22	56	21	53	12·0
								30	70	28	65	25	62	23	58	22	55	12·5
								33	71	30	67	27	63	25	60	23	56	13·0
										31	69	29	65	26	61	25	58	13·5
										33	70	30	66	28	63	26	59	14·0
												32	68	29	65	27	61	14·5
												34	70	31	66	29	63	15
														34	69	32	65	16
																35	68	17

Note: the pairs of figures are lower and upper confidence limits for p, given to two places of decimals.

Table 1 (*cont.*)

$n = 34$		$n = 36$		$n = 38$		$n = 40$		$n = 42$		$n = 44$		$n = 46$		$n = 48$		$n = 50$		$X+Y$
00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	0
00	05	00	05	00	05	00	04	00	04	00	04	00	04	00	04	00	03	0·2
00	07	00	07	00	06	00	06	00	06	00	06	00	05	00	05	00	05	·4
00	08	00	08	00	08	00	07	00	07	00	07	00	06	00	06	00	06	·6
00	09	00	09	00	08	00	08	00	08	00	07	00	07	00	07	00	06	·8
00	10	00	09	00	09	00	08	00	08	00	08	00	07	00	07	00	07	1·0
00	12	00	11	00	10	00	10	00	09	00	09	00	09	00	08	00	08	1·2
00	13	00	12	00	11	00	11	00	10	00	10	00	10	00	09	00	09	1·4
00	14	00	13	00	12	00	12	00	11	00	11	00	10	00	10	00	09	1·6
00	14	00	14	00	13	00	12	00	12	00	11	00	11	00	10	00	10	1·8
00	15	00	14	00	13	00	13	00	12	00	12	00	11	00	11	00	10	2·0
00	18	00	17	00	16	00	15	00	15	00	14	00	13	00	13	00	12	2·5
01	19	01	18	01	17	01	17	01	16	01	15	01	15	01	14	01	13	3·0
01	22	01	21	01	20	01	19	01	18	01	17	01	16	01	16	01	15	3·5
03	23	02	22	02	21	02	20	02	19	02	18	02	18	02	17	02	16	4·0
03	26	03	24	03	23	03	22	02	21	02	20	02	19	02	19	02	18	4·5
04	27	04	26	04	25	04	23	03	22	03	21	03	21	03	20	03	19	5·0
05	29	05	28	04	27	04	25	04	24	04	23	04	22	03	21	03	21	5·5
06	31	06	29	05	28	05	27	05	25	05	24	04	23	04	22	04	22	6·0
07	33	07	31	06	30	06	28	06	27	05	26	05	25	05	24	05	23	6·5
08	34	08	33	07	31	07	30	07	28	06	27	06	26	06	25	05	24	7·0
09	36	09	35	08	33	08	31	07	30	07	29	07	28	06	27	06	26	7·5
10	38	10	36	09	34	09	33	08	31	08	30	08	29	07	28	07	27	8·0
11	40	10	38	10	36	09	34	09	33	08	31	08	30	08	29	07	28	8·5
12	41	12	39	11	37	10	36	10	34	09	33	09	31	09	30	08	29	9·0
13	43	13	41	12	39	11	37	11	36	10	34	10	33	09	31	09	30	9·5
15	44	14	42	13	40	12	38	12	37	11	35	11	34	10	33	10	31	10·0
16	46	15	44	14	42	13	40	12	38	12	37	11	35	11	34	10	33	10·5
17	47	16	45	15	43	14	41	14	39	13	38	12	36	12	35	11	34	11·0
18	49	17	47	16	45	15	43	14	41	14	39	13	38	12	36	12	35	11·5
19	51	18	48	17	46	16	44	15	42	15	40	14	39	13	37	13	36	12·0
20	52	19	50	18	47	17	45	16	43	15	42	15	40	14	39	13	37	12·5
22	54	21	51	19	49	18	47	17	45	16	43	16	41	15	40	14	38	13·0
23	55	22	53	20	50	19	48	18	46	17	44	16	42	16	41	15	39	13·5
24	57	23	54	22	51	20	49	19	47	18	45	18	43	17	42	16	40	14·0
25	58	24	55	22	53	21	51	20	48	19	47	18	45	17	43	17	42	14·5
27	59	25	57	24	54	22	52	21	50	20	48	19	46	18	44	18	42	15·0
28	61	26	58	25	56	23	53	22	51	21	49	20	47	19	45	18	44	15·5
30	62	28	59	26	57	25	54	23	52	22	50	21	48	20	46	19	45	16·0
31	64	29	61	27	58	26	56	24	53	23	51	22	49	21	47	20	46	16·5
32	65	30	62	28	59	27	57	25	54	24	52	23	50	22	48	21	47	17·0
33	67	31	64	29	61	28	58	26	56	25	54	24	52	23	50	22	48	17·5
35	68	33	65	31	62	29	59	28	57	26	55	25	52	24	51	23	49	18·0
		34	66	32	63	30	61	28	58	27	56	26	54	25	52	23	50	18·5
		35	67	33	64	31	62	30	59	28	57	27	55	26	53	24	51	19·0
				34	66	32	63	31	60	29	58	28	56	26	54	25	52	19·5
				36	67	34	64	32	61	30	59	29	57	27	55	26	53	20·0
						35	65	33	63	31	60	30	58	28	56	27	54	20·5
						36	66	34	64	32	61	31	59	29	57	28	55	21·0
								35	65	33	62	32	60	30	58	29	56	21·5
								36	66	34	63	33	61	31	59	30	57	22·0
										35	65	34	62	32	60	31	58	22·5
										37	66	35	63	33	61	32	59	23
												37	65	35	63	34	61	24
														37	65	35	63	25
																37	65	26

Note: the pairs of figures are lower and upper confidence limits for p, given to two places of decimals.

Table 2. *Neyman-shortest unbiased 0·99-confidence intervals for p based on X, binomial (n, p) and an independent Y, rectangular (0, 1)*

$n=2$		$n=3$		$n=4$		$n=5$		$n=6$		$n=7$		$n=8$		$n=9$		$n=10$		$X+Y$
00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	0
00	90	00	68	00	54	00	44	00	37	00	33	00	29	00	26	00	24	0·1
00	95	00	78	00	63	00	53	00	45	00	40	00	36	00	32	00	30	·2
00	97	00	82	00	68	00	57	00	49	00	44	00	39	00	36	00	33	·3
00	98	00	84	00	71	00	60	00	52	00	46	00	42	00	38	00	35	·4
00	98	00	86	00	73	00	62	00	54	00	48	00	43	00	39	00	36	·5
00	98	00	87	00	74	00	64	00	56	00	49	00	45	00	41	00	38	·6
00	99	00	88	00	76	00	65	00	57	00	51	00	46	00	42	00	39	·7
00	99	00	89	00	77	00	67	00	58	00	52	00	47	00	43	00	39	·8
00	99	00	89	00	78	00	68	00	59	00	53	00	48	00	44	00	40	·9
00	99	00	90	00	78	00	68	00	60	00	54	00	49	00	44	00	41	1·0
00	1·0	00	96	00	84	00	74	00	65	00	58	00	52	00	48	00	44	1·1
00	1·0	00	98	00	88	00	77	00	68	00	61	00	55	00	51	00	47	1·2
00	1·0	00	98	00	90	00	79	00	71	00	63	00	57	00	52	00	48	1·3
00	1·0	00	99	00	91	00	81	00	72	00	65	00	59	00	54	00	50	1·4
00	1·0	00	99	00	92	00	82	00	74	00	66	00	60	00	55	00	51	1·5
		00	99	00	92	00	83	00	75	00	67	00	61	00	56	00	52	1·6
		00	99	00	93	00	84	00	76	00	68	00	62	00	57	00	53	1·7
		00	99	00	93	00	85	00	77	00	69	00	63	00	58	00	54	1·8
		00	99	00	94	00	85	00	77	00	70	00	64	00	58	00	54	1·9
		01	99	00	94	00	86	00	78	00	71	00	64	00	59	00	55	2·0
				00	97	00	89	00	81	00	74	00	67	00	62	00	57	2·1
				00	98	00	91	00	83	00	76	00	69	00	64	00	59	2·2
				00	99	00	93	00	85	00	77	00	71	00	65	00	60	2·3
				01	99	00	93	00	86	00	79	00	72	00	66	00	61	2·4
				01	99	01	94	00	87	00	80	00	73	00	67	00	62	2·5
						01	95	00	88	00	80	00	74	00	68	00	63	2·6
						01	95	01	88	01	81	00	75	00	69	00	64	2·7
						01	95	01	89	01	82	01	75	01	70	01	65	2·8
						02	96	02	89	01	82	01	76	01	70	01	65	2·9
						04	96	03	89	03	83	02	76	02	71	02	66	3·0
								03	92	03	85	02	79	02	73	02	68	3·1
								04	93	03	87	03	80	02	74	02	69	3·2
								04	94	03	88	03	81	02	76	02	70	3·3
								04	95	03	89	03	82	03	77	02	71	3·4
								05	95	04	89	03	83	03	77	02	72	3·5
										04	90	04	84	03	78	03	73	3·6
										05	90	04	84	03	79	03	74	3·7
										05	91	05	85	04	79	04	74	3·8
										07	91	06	85	05	80	04	75	3·9
										08	92	07	86	06	80	05	75	4·0
												07	88	06	82	06	77	4·1
												08	89	07	83	06	78	4·2
												08	90	07	84	06	79	4·3
												08	91	07	85	06	80	4·4
												09	91	08	86	07	81	4·5
														08	86	07	81	4·6
														09	87	08	82	4·7
														10	87	08	82	4·8
														11	88	09	83	4·9
														12	88	11	83	5·0
																11	84	5·1
																11	85	5·2
																11	86	5·3
																12	87	5·4
																12	88	5·5

Note: the pairs of figures are lower and upper confidence limits for p, given to two places of decimals.

Table 2 (*cont.*)

$n = 11$		$n = 12$		$n = 13$		$n = 14$		$n = 15$		$n = 16$		$n = 17$		$n = 18$		$n = 19$		$X+Y$
00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	0
00	22	00	20	00	19	00	18	00	17	00	16	00	15	00	14	00	13	0·1
00	27	00	25	00	23	00	22	00	21	00	19	00	18	00	17	00	17	·2
00	30	00	28	00	26	00	24	00	23	00	22	00	20	00	19	00	18	·3
00	32	00	30	00	28	00	26	00	24	00	23	00	22	00	21	00	20	·4
00	34	00	31	00	29	00	27	00	26	00	24	00	23	00	22	00	21	·5
00	35	00	32	00	30	00	28	00	27	00	25	00	24	00	23	00	22	·6
00	36	00	33	00	31	00	29	00	27	00	26	00	25	00	23	00	22	·7
00	37	00	34	00	32	00	30	00	28	00	27	00	25	00	24	00	23	·8
00	37	00	35	00	32	00	31	00	29	00	27	00	26	00	25	00	23	·9
00	38	00	35	00	33	00	31	00	29	00	28	00	26	00	25	00	24	1·0
00	43	00	40	00	38	00	35	00	33	00	32	00	30	00	29	00	27	1·2
00	46	00	43	00	41	00	38	00	36	00	34	00	32	00	31	00	29	1·4
00	48	00	45	00	42	00	40	00	38	00	36	00	34	00	32	00	31	1·6
00	50	00	47	00	44	00	41	00	39	00	37	00	35	00	33	00	32	1·8
00	51	00	48	00	45	00	42	00	40	00	38	00	36	00	34	00	33	2·0
00	55	00	51	00	48	00	46	00	43	00	41	00	39	00	37	00	35	2·2
00	57	00	54	00	51	00	48	00	45	00	43	00	41	00	39	00	37	2·4
00	59	00	55	00	52	00	49	00	47	00	44	00	42	00	40	00	38	2·6
00	60	00	57	00	53	00	50	00	48	00	45	00	43	00	41	00	39	2·8
02	61	01	58	01	54	01	51	01	49	01	46	01	44	01	42	01	40	3·0
02	64	02	61	01	57	01	54	01	51	01	49	01	46	01	44	01	42	3·2
02	67	02	63	02	59	02	56	01	53	01	50	01	48	01	46	01	44	3·4
02	68	02	64	02	61	02	57	02	54	02	52	01	49	01	47	01	45	3·6
03	69	03	65	03	62	02	58	02	55	02	53	02	50	02	48	02	46	3·8
05	70	04	66	04	63	04	59	03	56	03	54	03	51	03	49	03	47	4·0
05	73	05	69	04	65	04	62	04	59	03	56	03	53	03	51	03	49	4·2
06	75	05	71	05	67	04	63	04	60	04	57	03	55	03	52	03	50	4·4
06	76	06	72	05	68	05	65	04	61	04	59	04	56	04	53	03	51	4·6
07	77	07	73	06	69	06	66	05	62	05	59	04	57	04	54	04	52	4·8
09	78	08	74	08	70	07	66	06	63	06	60	06	58	05	55	05	53	5·0
10	81	09	76	08	72	07	68	07	65	06	62	06	59	05	57	05	55	5·2
11	82	09	78	09	74	08	70	07	67	07	64	06	61	06	58	06	56	5·4
11	83	10	79	09	75	09	71	08	68	07	65	07	62	06	59	06	57	5·6
13	84	11	80	10	76	10	72	09	69	08	66	08	63	07	60	07	58	5·8
15	85	13	81	12	77	11	73	10	69	09	66	09	63	08	61	08	58	6·0
		14	83	13	79	12	75	11	71	10	68	09	65	09	63	08	60	6·2
		15	84	13	80	12	76	11	73	10	69	10	66	09	64	08	61	6·4
		16	85	14	81	13	77	12	74	11	70	10	67	10	65	09	62	6·6
				16	82	14	78	13	74	12	71	11	68	10	65	10	63	6·8
				17	83	16	79	15	75	13	72	13	69	12	66	11	64	7·0
						16	81	15	77	14	74	13	71	12	68	11	65	7·2
						17	82	16	78	15	75	14	72	13	69	12	66	7·4
						18	83	17	79	15	76	14	73	13	70	12	67	7·6
								18	80	16	77	15	73	14	71	13	68	7·8
								19	81	18	77	17	74	16	71	15	68	8·0
										19	79	17	76	16	73	15	70	8·2
										19	80	18	77	17	74	16	71	8·4
										20	81	19	78	17	75	16	72	8·6
												20	78	18	75	17	73	8·8
												21	79	20	76	18	73	9·0
														20	77	19	74	9·2
														21	78	20	75	9·4
														22	79	20	76	9·6
																21	77	9·8
																23	77	10·0

Note: the pairs of figures are lower and upper confidence limits for p, given to two places of decimals.

Table 2 (*cont.*). 0·99 *confidence intervals*

$n = 20$		$n = 21$		$n = 22$		$n = 23$		$n = 24$		$n = 26$		$n = 28$		$n = 30$		$n = 32$		$X+Y$
00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	0
00	13	00	12	00	12	00	11	00	11	00	10	00	09	00	09	00	08	0·1
00	16	00	15	00	14	00	14	00	13	00	12	00	12	00	11	00	10	·2
00	18	00	17	00	16	00	16	00	15	00	14	00	13	00	12	00	11	·3
00	19	00	18	00	17	00	17	00	16	00	15	00	14	00	13	00	12	·4
00	20	00	19	00	18	00	17	00	17	00	16	00	15	00	14	00	13	·5
00	21	00	20	00	19	00	18	00	17	00	16	00	15	00	14	00	13	·6
00	21	00	20	00	20	00	19	00	18	00	17	00	16	00	15	00	14	·7
00	22	00	21	00	20	00	19	00	19	00	17	00	16	00	15	00	14	·8
00	22	00	21	00	21	00	20	00	19	00	18	00	16	00	15	00	15	·9
00	23	00	22	00	21	00	20	00	19	00	18	00	17	00	16	00	15	1·0
00	26	00	25	00	24	00	23	00	22	00	21	00	19	00	18	00	17	1·2
00	28	00	27	00	26	00	25	00	24	00	22	00	21	00	20	00	18	1·4
00	29	00	28	00	27	00	26	00	25	00	23	00	22	00	21	00	19	1·6
00	31	00	29	00	28	00	27	00	26	00	24	00	23	00	21	00	20	1·8
00	31	00	30	00	29	00	28	00	27	00	25	00	23	00	22	00	21	2·0
00	34	00	32	00	31	00	30	00	29	00	27	00	25	00	24	00	22	2·2
00	36	00	34	00	33	00	32	00	30	00	28	00	27	00	25	00	24	2·4
00	37	00	35	00	34	00	33	00	31	00	29	00	27	00	26	00	24	2·6
00	38	00	36	00	35	00	34	00	32	00	30	00	28	00	27	00	25	2·8
01	39	01	37	01	36	01	34	01	33	01	31	01	29	01	27	00	26	3·0
01	41	01	39	01	37	01	36	01	35	01	32	01	30	01	29	01	27	3·2
01	42	01	40	01	39	01	37	01	36	01	34	01	32	01	30	01	28	3·4
01	43	01	42	01	40	01	38	01	37	01	35	01	32	01	31	01	29	3·6
02	44	02	42	01	41	01	39	01	38	01	35	01	33	01	31	01	30	3·8
02	45	02	43	02	42	02	40	02	39	02	36	02	34	02	32	01	30	4·0
03	47	02	45	02	43	02	42	02	40	02	38	02	35	02	33	02	31	4·2
03	48	03	46	03	44	02	43	02	41	02	39	02	36	02	34	02	32	4·4
03	49	03	47	03	46	03	44	03	42	02	40	02	37	02	35	02	33	4·6
04	51	04	48	03	46	03	45	03	43	03	40	03	38	02	36	02	34	4·8
05	51	04	49	04	47	04	45	04	44	03	41	03	38	03	36	03	34	5·0
05	52	05	50	04	49	04	47	04	45	04	42	03	40	03	37	03	35	5·2
05	54	05	52	05	50	04	48	04	46	04	43	04	41	03	38	03	36	5·4
06	55	05	53	05	51	05	49	05	47	04	44	04	41	04	39	03	37	5·6
06	55	06	53	06	51	05	50	05	48	05	45	04	42	04	40	04	38	5·8
07	56	07	54	07	52	06	50	06	49	05	45	05	43	05	40	04	38	6·0
08	59	08	57	07	55	07	53	07	51	06	48	06	45	05	43	05	40	6·5
10	61	10	59	09	57	09	55	08	53	08	50	07	47	07	44	06	42	7·0
11	64	11	62	10	60	10	58	09	56	08	52	08	49	07	47	07	44	7·5
14	66	13	64	12	61	12	59	11	57	10	54	09	51	09	48	08	45	8·0
15	69	14	66	13	64	13	62	12	60	11	56	10	53	09	50	09	48	8·5
17	70	16	68	15	66	15	64	14	62	13	58	12	55	11	52	10	49	9·0
19	73	18	71	17	68	16	66	15	64	14	60	13	57	12	54	11	51	9·5
21	75	20	72	19	70	18	68	17	66	16	62	14	58	13	55	12	52	10·0
23	77	21	75	20	72	19	70	18	68	17	64	15	60	14	57	13	54	10·5
		24	76	22	74	21	72	20	69	18	65	17	62	16	58	15	56	11·0
				24	76	23	74	21	72	19	68	18	64	17	60	15	57	11·5
						25	75	24	73	21	69	20	65	18	62	17	59	12·0
								25	75	23	71	21	67	19	64	18	61	12·5
								27	76	25	72	22	68	21	65	19	62	13·0
										26	74	24	70	22	67	20	64	13·5
										28	75	25	72	23	68	22	65	14·0
												27	73	24	70	23	66	14·5
												28	75	26	71	24	68	15
														29	74	27	70	16
																30	73	17

Note: the pairs of figures are lower and upper confidence limits for p, given to two places of decimals.

Table 2 (*cont.*)

$n = 34$		$n = 36$		$n = 38$		$n = 40$		$n = 42$		$n = 44$		$n = 46$		$n = 48$		$n = 50$		$X+Y$
00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	0
00	10	00	09	00	09	00	08	00	08	00	07	00	07	00	07	00	07	·2
00	12	00	11	00	10	00	10	00	09	00	09	00	09	00	08	00	08	·4
00	13	00	12	00	11	00	11	00	10	00	10	00	09	00	09	00	09	·6
00	13	00	13	00	12	00	12	00	11	00	11	00	10	00	10	00	09	·8
00	14	00	13	00	13	00	12	00	12	00	11	00	11	00	10	00	10	1·0
00	16	00	15	00	15	00	14	00	13	00	13	00	12	00	12	00	11	1·2
00	17	00	16	00	16	00	15	00	14	00	14	00	13	00	13	00	12	1·4
00	18	00	17	00	17	00	16	00	15	00	14	00	14	00	13	00	13	1·6
00	19	00	18	00	17	00	16	00	16	00	15	00	14	00	14	00	13	1·8
00	20	00	19	00	18	00	17	00	16	00	15	00	15	00	14	00	14	2·0
00	23	00	22	00	20	00	20	00	19	00	18	00	17	00	16	00	16	2·5
00	24	00	23	00	22	00	21	00	20	00	19	00	18	00	18	00	17	3·0
01	27	01	26	01	24	01	23	01	22	00	21	00	20	00	20	00	19	3·5
01	29	01	27	01	26	01	25	01	24	01	23	01	22	01	21	01	20	4·0
02	31	02	29	01	28	01	27	01	26	01	25	01	24	01	23	01	22	4·5
03	32	02	31	02	29	02	28	02	27	02	26	02	25	02	24	02	23	5·0
03	35	03	33	03	32	03	30	02	29	02	28	02	27	02	26	02	25	5·5
04	36	04	34	04	33	03	31	03	30	03	29	03	28	03	27	03	26	6·0
05	38	04	37	04	35	04	33	04	32	03	31	03	29	03	28	03	27	6·5
06	40	05	38	05	36	05	35	05	33	04	32	04	30	04	29	04	28	7·0
06	42	06	40	06	38	05	36	05	35	05	33	05	32	04	31	04	30	7·5
07	43	07	41	07	39	06	38	06	36	06	35	05	33	05	32	05	31	8·0
08	45	08	43	07	41	07	39	07	38	06	36	06	35	06	33	05	32	8·5
09	47	09	44	08	42	08	40	07	39	07	37	07	36	06	34	06	33	9·0
10	48	10	46	09	44	08	42	08	40	08	39	07	37	07	36	07	35	9·5
11	50	11	47	10	45	10	43	09	42	09	40	08	38	08	37	08	36	10·0
12	52	11	49	11	47	10	45	10	43	09	41	09	40	08	38	08	37	10·5
14	53	13	50	12	48	11	46	11	44	10	42	10	41	09	39	09	38	11·0
14	55	14	52	13	50	12	48	11	46	11	44	10	42	10	41	09	39	11·5
16	56	15	53	14	51	13	49	12	47	12	45	11	43	11	42	10	40	12·0
17	58	15	55	15	53	14	50	13	48	12	46	12	45	11	43	11	42	12·5
18	59	17	56	16	54	15	51	14	49	13	47	13	46	12	44	12	43	13·0
19	61	18	58	17	55	16	53	15	51	14	49	13	47	13	45	12	44	13·5
20	62	19	59	18	56	17	54	16	52	15	50	14	48	14	46	13	45	14·0
21	63	20	61	19	58	18	56	17	53	16	51	15	49	14	48	14	46	14·5
23	64	21	62	20	59	19	57	18	54	17	52	16	50	15	49	15	47	15·0
24	66	22	63	21	61	20	58	19	56	18	54	17	52	16	50	15	48	15·5
25	67	23	64	22	62	21	59	20	57	19	55	18	53	17	51	16	49	16·0
26	69	24	66	23	63	22	61	20	58	19	56	18	54	18	52	17	50	16·5
28	70	26	67	24	64	23	61	22	59	20	57	19	55	19	53	18	51	17·0
29	71	27	68	25	66	24	63	22	60	21	58	20	56	19	54	18	52	17·5
30	72	28	69	26	67	25	64	24	61	22	59	21	57	20	55	19	53	18·0
		29	71	27	68	26	65	24	63	23	60	22	58	21	56	20	54	18·5
		31	72	29	69	27	66	26	64	24	61	23	59	22	57	21	55	19·0
				30	70	28	68	26	65	25	63	24	60	23	58	22	56	19·5
				31	71	29	68	28	66	26	64	25	61	24	59	23	57	20·0
						30	70	28	67	27	65	26	62	24	60	23	58	20·5
						32	71	30	68	28	66	27	63	26	61	24	59	21·0
								31	69	29	67	28	65	26	62	25	60	21·5
								32	70	30	68	29	65	27	63	26	61	22·0
										31	69	29	67	28	64	27	62	22·5
										32	70	31	67	29	65	28	63	23
												33	69	31	67	30	65	24
														33	69	31	67	25
																33	69	26

Note: the pairs of figures are lower and upper confidence limits for p, given to two places of decimals.

Biometrika (1960), **47**, 3 *and* 4, *p.* 363
With 1 *text-figure*
Printed in Great Britain

On the power function of the exact test for the 2×2 contingency table

BY B. M. BENNETT AND P. HSU
University of Washington, Seattle

1. *Introduction.* For the 2×2 table which consists of $N(= A+B)$ individuals in two groups

Group	With attribute	Without attribute	Total
1	a	$A-a$	A
2	b	$B-b$	B
Total	$r = a+b$	$N-r$	N

classified according to whether or not they possess a certain character or attribute C, an exact test of significance for the hypothesis H_0: $p_1 = p_2$ of the equality of the respective population proportions of individuals possessing C was devised concurrently by Yates (1934) and Fisher (1935). Under the null hypothesis H_0 they demonstrated that the conditional probability that a of the individuals in the array above with fixed marginal totals r, A, B possess the attribute is given by the successive terms of the hypergeometric distribution

$$f(a\,|\,r) = \frac{\binom{A}{a}\binom{B}{b}}{\binom{N}{r}} = \frac{\binom{A}{a}\binom{B}{r-a}}{\binom{N}{r}}, \tag{1}$$

where $r = a+b$, $a = \max(0, r-B)$, $\max(0, r-B)+1, \ldots, \min(r, A)$, or $b = \max(0, r-A)$, $\max(0, r-A)+1, \ldots, \min(r, B)$. In terms of b we may write the distribution as

$$f(b\,|\,r) = \frac{\binom{A}{r-b}\binom{B}{b}}{\binom{N}{r}}. \tag{2}$$

For the exact test, the probabilities (1) are then to be summed over the given observed or more extreme values of a, or b in (2).

For small frequencies, significant values of b in particular have been provided, i.e. the critical values b_0 such that

$$P\{b \leqslant b_0\,|\,r\} = \sum_{b=0}^{b_0} f(b\,|\,r) \leqslant \alpha \quad \text{and} \quad P\{b \leqslant b_0+1\,|\,r\} > \alpha.$$

For nominal levels of significance $\alpha = 0{\cdot}05, 0{\cdot}025, 0{\cdot}01, 0{\cdot}005$, Finney (1948) has computed b_0 for $B \leqslant A \leqslant 15$, and Latscha (1953) has supplemented these when $B \leqslant A \leqslant 20$. The present authors have recently extended these exact probabilities to the cases $B \leqslant A \leqslant 25$. The particular arrangement of these tables is mainly concerned with one-sided alternatives: $p_1 > p_2$. For two-sided alternatives of the form: $p_1 \neq p_2$, the tabular exact probabilities are accordingly to be doubled.

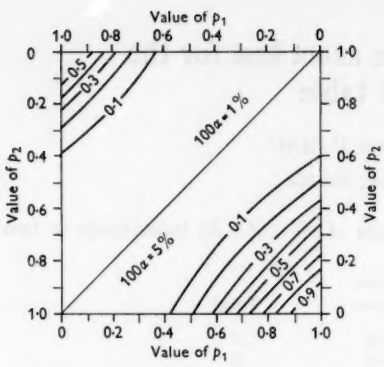


Fig. 1. Power contours ($A = 5$, $B = 5$).

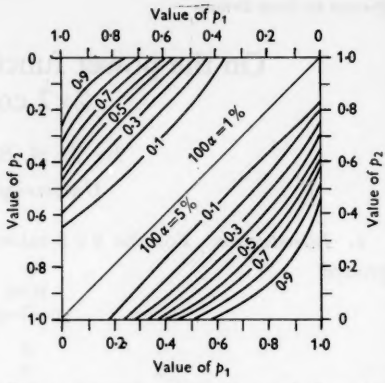


Fig. 2. Power contours ($A = 10$, $B = 10$).

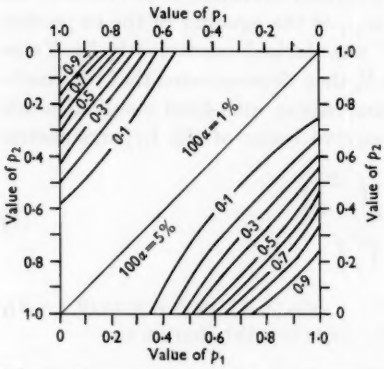


Fig. 3. Power contours ($A = 10$, $B = 5$).

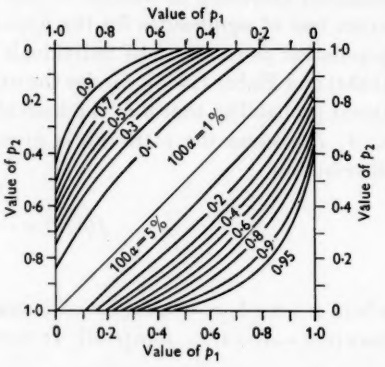


Fig. 4. Power contours ($A = 15$, $B = 15$)

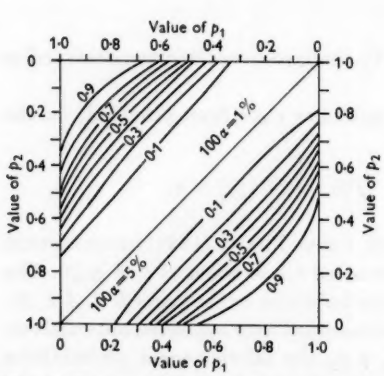


Fig. 5. Power contours ($A = 15$, $B = 10$).

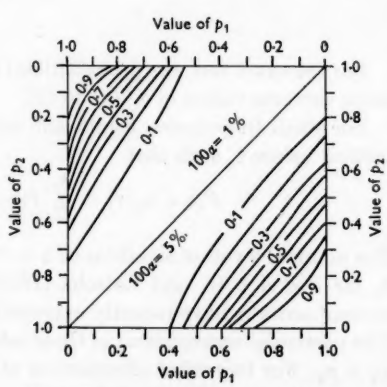


Fig. 6. Power contours ($A = 15$, $B = 5$).

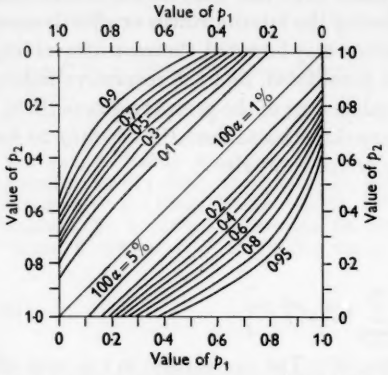


Fig. 7. Power contours ($A = 20$, $B = 20$).

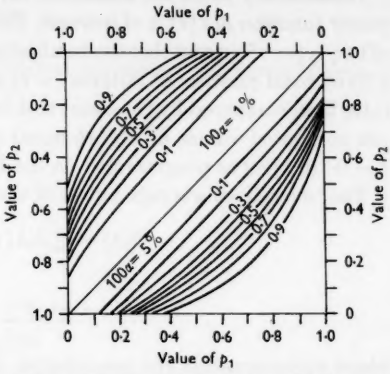


Fig. 8. Power contours ($A = 20$, $B = 15$).

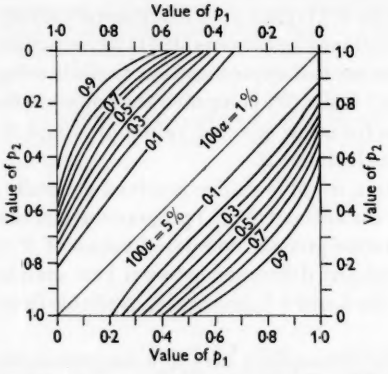


Fig. 9. Power contours ($A = 20$, $B = 10$).

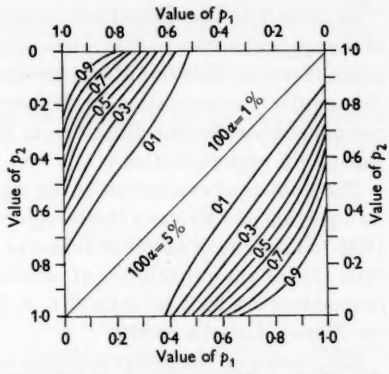


Fig. 10. Power contours ($A = 20$, $B = 5$).

2. *Conditional power function of exact test.* Under the alternative hypothesis it is easily verified that the conditional distribution of b, given $r = a+b$, is

$$g(b\,|\,r) = \frac{\binom{A}{r-b}\binom{B}{b}\lambda^b}{\sum_{i=m_1}^{m_2}\binom{A}{r-i}\binom{B}{i}\lambda^i} \tag{3}$$

where

$$\lambda = \left(\frac{p_2}{1-p_2}\right)\Big/\left(\frac{p_1}{1-p_1}\right)$$

represents the probability ratio, or relative risk, and $m_1 = \max(0, r-A)$, $m_2 = \max(r, B)$. The conditional power function of the exact test is then

$$\beta(\lambda\,|\,r) = P(b \leqslant b_0\,|\,H) = \sum_{b=0}^{b_0} g(b\,|\,r) = \sum_{b=0}^{b_0}\binom{A}{r-b}\binom{B}{b}\lambda^b \Big/ \sum_{i=m_1}^{m_2}\binom{A}{r-i}\binom{B}{i}\lambda^i \tag{4}$$

or the ratio of two polynomials in λ. The probability distribution (4) was first developed by Fisher (1935) in the determination of fiducial limits for the ratio λ.

Occasionally there are situations involving the exact 2×2 test wherein the conditional power function $\beta(\lambda \mid r)$ is of interest. Thus in considering the interim status or effectiveness of two types of current therapy in which success or failure may be noted, those cases involving a fixed total number of failures ($= r$) may first be considered, and the respective failure rates then compared by the exact test for the 2×2 table. One of the present authors (P.H.) has computed values of $\beta(\lambda \mid r)$ based on the Finney–Latscha tables: $B \leqslant A$, 5(5) 20 for $\lambda = 0{\cdot}1(0{\cdot}2)0{\cdot}9$ as programmed on the IBM-650 electronic computer.

The 'overall' or average power of the exact test is

$$\beta(\lambda) = \sum_r \beta(\lambda \mid r)\, p(r)$$

$$= q_1^A q_2^B \sum_r \left(\frac{p_1}{1-p_1}\right)^r \left[\sum_{b=m_1}^{b_0} C_{r-b}^A C_b^B \lambda^b\right], \tag{5}$$

where $p(r)$ represents the convolution or distribution of r. The summation in r is over all significant pairs of points on the diagonals, $r = a + b$, in the (a, b) sample space at level of significance equal to α at most, as determined from the Finney–Latscha tables.

Patnaik (1948) developed normal approximations for $\beta(\lambda \mid r)$ and $\beta(\lambda)$. The general validity of the approximation to $\beta(\lambda \mid r)$ was subsequently questioned by Stevens (1951) who supplied a modification. Sillitto (1949) provided still another normal approximation to $\beta(\lambda)$, using the angular transformation for the binomial variates a and b. He computed the power of the normalized test for the difference in the two angles for a few selected values of A and B, though the approximation is not very precise for small A and B.

The numerical evaluation of the double summation in (5) is rather involved generally. As a portion of a Master's thesis requirement one of the authors (P.H.) programmed on the IBM-704 a series of tables of values of $\beta(\lambda)$. These tables give the computed values of $\beta(\lambda)$ from (5), or the probability of establishing a significant difference between two sample proportions in a 2×2 table for $B \leqslant A$, 5(5)20, using the 5 and 1 % levels of b_0 available from the Finney–Latscha results.

The results on the power function calculations are summarized here in a diagrammatic representation (Figs. 1–10). Contours of equal power in the (p_1, p_2) plane, prepared by graphical interpolation, are shown for the single-tailed test of the hypothesis: $p_1 = p_2$ against the alternative: $p_1 > p_2$, for the various combinations of A and B and for the 5 and 1 % levels of significance. The contours for the 5 % level test are drawn below the diagonal and that of the 1 % level above the diagonal with the scales in p_1 and p_2 reversed. It is clear that when the value of B is much smaller than that of A, an unsymmetric series of contours results.

As regards the accuracy of Sillitto's approximation (1949), the following comparisons were made:

					Power function	
A	B	p_1	p_2	α	$\beta(\lambda)$	Sillitto approx.
10	10	0·8	0·2	0·05	0·8054	0·891
20	15	0·8	0·2	0·05	0·9681	0·983

The authors would like to acknowledge the assistance of the Research Computer Laboratory of the University of Washington, and also certain financial support from the State of Washington, Initiative 171 Fund for Research in Biology and Medicine.

References

FINNEY, D. J. (1948). *Biometrika*, **35**, 145–56.

FISHER, R. A. (1934). *Statistical Methods for Research Workers*, 5th edition. Edinburgh: Oliver and Boyd.

FISHER, R. A. (1935). *J. R. Statist. Soc.* **98**, 39–54.

HSU, P. (1960). Master's Thesis, University of Washington, (unpublished).

LATSCHA, R. (1955). *Biometrika*, **40**, 74–86.

PATNAIK, P. B. (1948). *Biometrika*, **35**, 157–75.

SILLITTO, G. P. (1949). *Biometrika*, **36**, 347–52.

STEVENS, W. L. (1951). *Biometrika*, **38**, 468–72.

YATES, F. (1934). *J. R. Statist. Soc.* Suppl. **1**, 217–35.

Editorial note

The authors of the preceding paper have referred to the papers by Patnaik (1948) and Sillitto (1949) in which approximate methods were suggested for obtaining the overall power function of the tests for the significance of a difference between two binomial probabilities, p_1 and p_2. In the two comparisons given at the end of their paper, it is seen that Sillitto's approximation gives considerably higher power values than those of the present paper. This tendency is also apparent for other values of A and B. For example, Sillitto (1949, p. 350) in his Table 2(a) gives some comparisons for the case $A = B = 15$. These are for the two-tailed test with $\alpha = 0{\cdot}10$ and $0{\cdot}02$, but as p_1 and p_2 draw apart, these results will scarcely differ from those for a single-tailed test with $\alpha = 0{\cdot}05$ and $0{\cdot}01$.

By combining some entries from Sillitto's Table 2(a), from Patnaik's Table 3(a) and from the original Bennett & Hsu calculations on which their power contours of Fig. 4 were based,* the following table can be made up:

Power of tests. Case $A = B = 15$

		Nominal significance level ($\alpha = 0{\cdot}05$)				Nominal significance level ($\alpha = 0{\cdot}01$)			
		Patnaik				Patnaik			
p_1	p_2	'Exact' values	2nd approx.	Sillitto approx.	Bennett & Hsu	'Exact' values	2nd approx.	Sillitto approx.	Bennett & Hsu
0·3	0·1	0·389	0·422	0·409	0·256	0·149	0·193	0·180	0·076
·6	·1	·919	·953	·926	·845	·739	·844	·778	·609
·7	·2	·896	·912	·893	·807	·680	·750	·713	·527
·8	·2	·974	·982	·970	·939	·872	·923	·885	·765

At first sight it seems anomalous that the 'Exact values' of Patnaik are all so much larger than the values of Bennett & Hsu. On examination of Patnaik's paper, however, it seems clear that what he calculated was the sum of the products of the appropriate binomial probabilities associated with points in the (a, b) lattice falling beyond the significance levels obtained by assuming that the hypergeometric distributions of the null hypothesis could be represented by continuous normal distributions. Thus Patnaik's power values were only 'exact' for the particular significance points resulting from this approximation.

In the same way Sillitto's approximation, using the angular transformation is based on assuming that the transformed variable is normally distributed. Bennett & Hsu have, however, used Finney's (1948) significance points for 'b given r' and if Finney's table is examined it is seen that the sum of the tail terms of the hypergeometric series are, owing to discontinuity, almost always considerably less than the nominal 0·05 and 0·01 values. The effective 'size' of the test based on Finney's table is probably, therefore, considerably less than that involved in the test using either approximation and, as a natural result, the power is less.

* These results were attached to an earlier revision of their paper.

Whereas, however, we can make a shrewd guess at the 'size' of the test using Finney's tables, it is impossible to be sure what the 'size' will be for the approximate tests. For samples larger than 20 the Finney–Latscha form of table is not of course at present available and, no doubt, as A and B increase both the size and power obtained from the approximations will approach their nominal values.

An interesting point which the Bennett & Hsu charts bring out very clearly is how difficult it is to distinguish between two widely different probabilities in small samples. For example, even when $A = B = 20$, the odds are only just in favour of detecting a difference between a $p_1 = 0{\cdot}5$ and a $p_2 = 0{\cdot}2$, using a nominal 5 % significance level (power = 0·53). E.S.P.

Biometrika (1960), **47**, 3 *and* 4, *p.* 399
With 1 *text-figure*
Printed in Great Britain

Table of the bounds of the probability integral when the first four moments are given

BY J. A. SIMPSON AND B. L. WELCH
University of Leeds

1. INTRODUCTION

The calculation of upper and lower bounds for the probability integral, $\Pr(x \leqslant \alpha)$, of a distribution whose first m moments, $\theta_1, \theta_2, \ldots, \theta_m$, are given, is a problem for which a considerable general theory has long existed (for earlier references see, for instance, Uspensky, 1937). The present paper is concerned with the tabulation of the bounds when $m = 4$, and with the provision of some commentary on the tabular values. A main reason for undertaking this investigation is that we wished to provide some concrete figures which may be useful in teaching in connexion with the representation of frequency distributions by moments. Our tabulation shows what gains over the more elementary Tchebychef inequalities (cf. equation (4), below) are obtainable as soon as the skewness and kurtosis of a distribution are specified. A table of triple entry is necessary for the presentation of our results and accordingly, to cover the full range of the parameters involved, we have been forced to tabulate on a very broad net. (For a previous table on a finer net of values of α, but covering only a very limited range of all three parameters, see Berge (1932.) Other tables of varying extent must undoubtedly exist.)

A brief summary of the formulae used in computing the table is provided, since direct calculation from these formulae will often be the best way of interpolating values not tabulated. In § 4 some discussion is also given of another problem which may be made to depend on the one for which the tabulation is explicitly done. This arises when the range of x is limited, say to (b_1, b_2), but where now only the skewness of the distribution is known in addition to the first two moments. This will naturally include the case where one of the limits, b_1 or b_2, is at infinity.

2. VARIABLE UNRESTRICTED IN RANGE

If all our information about the probability distribution function $dP(x)$ lies in the values of the first four moments, it may be shown that the calculation of bounds for $\Pr(x \leqslant \alpha)$ depends on the following preliminaries.

Let us associate with the point $x = \alpha$ two further points $x = \rho_1$ and $x = \rho_2$ and define a 3-point probability distribution consisting of probabilities c on α, d_1 on ρ_1 and d_2 on ρ_2. Suppose that the five quantities ρ_1, ρ_2, c, d_1 and d_2 which we have introduced are then chosen so that the resulting 3-point probability distribution is equimomental with $\mathrm{d}P(x)$ as far as the fourth moment. If x (and α) are expressed in standard measure, so that θ_3 and θ_4 are equal to the usual measures of skewness and kurtosis $\sqrt{\beta_1}$ and β_2, respectively, it may be shown that ρ_1 and ρ_2 are roots of the quadratic equation

$$x^2(\alpha^2 - \alpha\theta_3 - 1) + x\{\alpha\Delta - \theta_3(\alpha^2 - \alpha\theta_3 - 1)\} + \{\Delta - (\alpha^2 - \alpha\theta_3 - 1)\} = 0, \qquad (1)$$

where we write $\Delta = \theta_4 - \theta_3^2 - 1 = \beta_2 - \beta_1 - 1,$

an essentially positive quantity.

Furthermore, we find that

$$\left.\begin{aligned} c &= (1+\rho_1\rho_2)(\alpha-\rho_1)^{-1}(\alpha-\rho_2)^{-1}, \\ d_1 &= (1+\rho_2\alpha)(\rho_1-\rho_2)^{-1}(\rho_1-\alpha)^{-1}, \\ d_2 &= (1+\rho_1\alpha)(\rho_2-\alpha)^{-1}(\rho_2-\rho_1)^{-1}. \end{aligned}\right\} \tag{2}$$

The roots ρ_1 and ρ_2 are always real and the values c, d_1 and d_2 always positive which means that we have defined a real probability distribution which is a possible (although extreme) form which the actual $dP(x)$ could take. The classical theory now asserts that this extreme distribution provides us with the bounds of $\Pr(x \leqslant \alpha)$. The lower bound is given by summing those values of d_i corresponding to the roots ρ_i (either 0, 1 or 2 in number) which fall below α. The upper bound is given by adding the value of c to this sum. If, for definiteness, we denote the least root by ρ_1 then, according to how α lies with respect to ρ_1 and ρ_2, we find

$$\left.\begin{aligned} 0 < \Pr(x \leqslant \alpha) &\leqslant c && \text{if} \quad \alpha \leqslant \lambda_1, \\ d_1 < \Pr(x \leqslant \alpha) &\leqslant c+d_1 && \text{if} \quad \lambda_1 \leqslant \alpha \leqslant \lambda_2, \\ d_1+d_2 < \Pr(x \leqslant \alpha) &\leqslant 1 && \text{if} \quad \lambda_2 \leqslant \alpha, \end{aligned}\right\} \tag{3}$$

where λ_1 and λ_2 are roots of the equation $x^2-\theta_3 x-1=0$, i.e.

$$\lambda_1 = \tfrac{1}{2}\{\theta_3-\sqrt{(\theta_3^2+4)}\}, \quad \lambda_2 = \tfrac{1}{2}\{\theta_3+\sqrt{(\theta_3^2+4)}\}.$$

The lower bounds in (3) are attainable to within as small a distance as we please.

3. COMMENTS ON TABLE 1

The quantities c, d_1 and d_2 in (2) are functions of α, θ_3 and θ_4 (or equivalently of α, θ_3 and Δ). A table of the solution (3) must therefore be of triple entry. In Table 1 we provide numerical values of the lower and upper bounds L and U when

$$\theta_3 = 0\,(1)\,4,\,8; \quad \alpha = -10\,(1)-3, -2{\cdot}5\,(0{\cdot}5)\,2{\cdot}5,\,3\,(1)\,10;$$

and

$$\Delta = 0, 0{\cdot}5, 1, 2, 3, 4, 6, 8, 12, 16, \infty;$$

in all combinations. The values for given θ_3 are isolated in separate subtables. (Values for θ_3 negative may be deduced by changing the sign of the variable at the outset and rephrasing the problem in an obvious way.)

This mode of presentation brings into prominence the fact that in general, for given θ_3 and α, the range (L, U) corresponding to any Δ completely overlaps that for any smaller Δ. At one extreme when $\Delta = 0$ we have $(U-L) = 0$ since in this case $dP(x)$ must be concentrated only on the two points λ_1 and λ_2 defined above, the probabilities associated with these points being $\lambda_2(\lambda_2-\lambda_1)^{-1}$ and $\lambda_1(\lambda_1-\lambda_2)^{-1}$, respectively. At the other extreme, when $\Delta \to \infty$, one of the roots of (1) goes to infinity whilst the other is $-\alpha^{-1}$. Equations (2) then lead to a probability $(1+\alpha^2)^{-1}$ associated with α, a probability $\alpha^2(1+\alpha^2)^{-1}$ at $-\alpha^{-1}$, together with an infinitesimal probability at an infinitely distant point. Hence, whatever θ_3, provided only that it is finite, we have when $\Delta \to \infty$:

$$\left.\begin{aligned} &\text{(i) for } \alpha < 0 && L = 0, && U = (1+\alpha^2)^{-1}; \\ &\text{(ii) for } \alpha > 0 && L = \alpha^2(1+\alpha^2)^{-1}, && U = 1. \end{aligned}\right\} \tag{4}$$

These bounds do not depend on θ_3 at all and indeed are simply the Tchebychef bounds for the integral when only the first two moments of the distribution are given. Hence knowledge of θ_3 does not narrow the bounds (L, U) unless it is accompanied by some restriction on Δ (or θ_4). When this further information is available then a pair of bounds will be obtained which are intermediate between those for $\Delta = 0$ and $\Delta = \infty$.

This behaviour of the solution is illustrated diagrammatically in Fig. 1, which relates to the case $\theta_3 = \sqrt{\beta_1} = 1$. The outer continuous curves here are the Tchebychef bounds of equation (4) which would be the same whatever θ_3. The step function $ABCDEF$ relates to the case $\Delta = 0$ when the distribution $dP(x)$ must be concentrated on the two points

$$\lambda_1 = \tfrac{1}{2}\{\theta_3 - \sqrt{(\theta_3^2+4)}\} = -0{\cdot}618 \quad \text{and} \quad \lambda_2 = \tfrac{1}{2}\{\theta_3 + \sqrt{(\theta_3^2+4)}\} = 1{\cdot}618,$$

so that for all values of α (excepting perhaps λ_1 and λ_2) we have $L = U$. The dashed curves (together with AB and EF) are the lower and upper bounds when $\Delta = 0{\cdot}5$. For any Δ

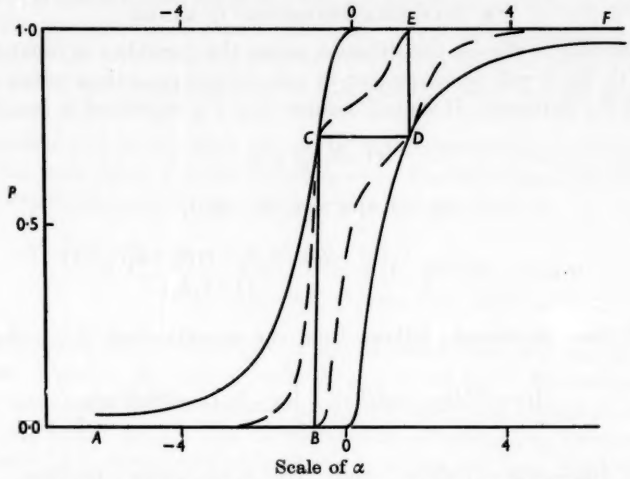


Fig. 1. Upper and lower bounds for $\Pr(x \leqslant \alpha)$ when $\theta_3 = \sqrt{\beta_1} = 1$. (i) $ABCDEF$ is the cumulative distribution function of x if $\Delta = 0$. (ii) The dashed curves (completed by AB and EF) are the lower and upper bounds when $\Delta = 0{\cdot}5$. (iii) The outer curves are the bounds when $\Delta = \infty$.

between 0·5 and ∞ the corresponding bounds will lie between the dashed and continuous curves. For any Δ between 0 and 0·5 the solution will lie between the step function and the dashed curves.

The points λ_1 and λ_2 which are determined by θ_3 and which are roots of the orthogonal polynomial of degree two, associated with $dP(x)$, obviously play an important part. For all values of Δ the curves coalesce at these points and hence for α in the vicinity of λ_1 and λ_2 knowledge of θ_3 and Δ is not enabling us in general to improve upon (4). As we depart from λ_1 and λ_2 either toward the middle or toward the tails of $dP(x)$ we begin to obtain narrower bounds. The tails are, perhaps, of most interest. If, in Table 1, we concentrate attention, for given θ_3, on a given value of α below λ_1 we see, from the values of U, how the maximum amount of probability in the tail below α falls to zero as Δ diminishes. On the other hand for given α exceeding λ_2 the values of L also show how the maximum amount in the upper tail above α falls to zero as Δ diminishes.

The improvement in our inequalities is therefore seen to be appreciable although, as has often been pointed out, the fact that the extreme bounds are provided by a 3-point distribution implies that greatly improved inequalities can be obtained if further information about the form of $dP(x)$ is available. (For a recent discussion of more restrictive problems the reader may be referred, for instance, to Mallows (1956).) Despite this, however, it is still of value, at least from a didactic point of view, to indicate numerically in a broad way, as is done in Table 1, just what are the gains which flow from the Tchebychef theory by using θ_3 and θ_4 in addition to θ_1 and θ_2. We all encounter the common measures of skewness and kurtosis at an early stage in our study of statistical methods, and it seems to be important to have exact information available over the whole range of the parameters concerned, showing how far these measures alone take us in the 'representation of frequency distributions'.

4. VARIABLE RESTRICTED IN RANGE

Before proceeding to discuss the situation where the variable x is restricted to lie in a known range (b_1, b_2) it will be convenient to note certain conditions which must connect b_1 and b_2 with the moments. If we still assume that x is expressed in standard measure, these are

$$(1+b_1b_2) \leqslant 0, \tag{5}$$

$$(b_1-b_1^{-1}) \leqslant \theta_3 \leqslant (b_2-b_2^{-1}), \tag{6}$$

$$0 \leqslant \Delta = (\theta_4-\theta_3^2-1) \leqslant \frac{-(b_1^2-b_1\theta_3-1)(b_2^2-b_2\theta_3-1)}{(1+b_1b_2)}. \tag{7}$$

The proof of these statements follows from the consideration that, whatever δ, the quantities

$$\int(x-b_1)(b_2-x)\,dP(x), \quad \int(x-b_1)(x-\delta)^2\,dP(x),$$

$$\int(b_2-x)(x-\delta)^2\,dP(x) \quad \text{and} \quad \int(x-b_1)(b_2-x)(x-\delta)^2\,dP(x)$$

are essentially positive. We shall have occasion to use (7) below.

The solution of the problem of evaluating the extreme bounds of $\Pr(x \leqslant \alpha)$ when b_1, b_2 and moments up to θ_4 are given will obviously still be given by the analysis of § 2 for those values of α which are such that the roots (ρ_1, ρ_2) of equation (1) turn out to lie within (b_1, b_2). However, for certain values of α the roots (ρ_1, ρ_2) will not be so contained and other formulae are needed. The whole solution is conveniently set out in a form suitable for computation by Zelen (1954) and we shall not say anything further about it here beyond noting that Table 1 is now no longer universally applicable. Instead, we shall turn to a somewhat simpler limited range problem which is of interest in the present context in that, whatever α, its solution can always be linked up with Table 1. This is the problem where, in addition to b_1 and b_2, only the moments up to θ_3 are given (and not θ_4). The calculation of bounds for $\Pr(x \leqslant \alpha)$ is in this case seen to depend on the following preliminaries.

Let b be one of the ends of the range of x (which end to be specified later). Let us associate with the points α and b a third point $x = \rho$ and define a 3-point probability distribution c on α, d on ρ and e on b. Suppose that the four quantities ρ, c, d and e are chosen so that the

resulting 3-point distribution is equimomental with $dP(x)$ as far as the third moment. This requires that

$$\rho = (\theta_3 - \alpha - b)(1 + \alpha b)^{-1}, \tag{8}$$

$$\left.\begin{aligned} c &= (1 + b\rho)(\alpha - b)^{-1}(\alpha - \rho)^{-1}, \\ d &= (1 + b\alpha)(\rho - \alpha)^{-1}(\rho - b)^{-1}, \\ e &= (1 + \rho\alpha)(b - \rho)^{-1}(b - \alpha)^{-1}. \end{aligned}\right\} \tag{9}$$

So far we have not said whether $b = b_1$ or $b = b_2$. The choice will now be made so that, *simultaneously*, we find $b_1 \leqslant \rho \leqslant b_2$ and also c, d and e all positive. For any particular α it transpires that this condition will in general be satisfied by setting b equal to one and one only of the alternative values b_1 or b_2. We have thus defined in a unique fashion, through (8) and (9), a 3-point distribution on (α, b, ρ) which is a possible form which $dP(x)$ could actually have. It is now a known result that this extreme form provides us with the bounds of $\Pr(x \leqslant \alpha)$. The procedure is to observe how many of the two quantities b and ρ are less than α. The sum of the corresponding weights will give the lower bound of $\Pr(x \leqslant \alpha)$. The upper bound is obtained by adding c on to this sum (for more explicit details cf. Zelen (1954)).

Direct calculation in this way from (8) and (9) is straightforward, but alternatively we may make a link with Table 1 in the following way. For the 3-point distribution (c, d, e) on (α, ρ, b) the fourth moment, θ_4', say, can be evaluated and leads to

$$\begin{aligned} \Delta' = (\theta_4' - \theta_3^2 - 1) &= -(1 + b\alpha)(1 + b\rho)(1 + \alpha\rho) \\ &= -(b^2 - \theta_3 b - 1)(\alpha^2 - \theta_3\alpha - 1)(1 + \alpha b)^{-1}. \end{aligned} \tag{10}$$

If we substitute $b = b_1$ and $b = b_2$ in (10) it can be shown that only one of the resulting Δ' lies between 0 and $-(b_1^2 - \theta_3 b_1 - 1)(b_2^2 - \theta_3 b_2 - 1)(1 + b_1 b_2)^{-1}$. This value of Δ' must correspond to that choice of b which leads through (8) and (9) to a real 3-point probability distribution comprised in (b_1, b_2). For, as we noted in (7) above, any distribution lying in (b_1, b_2) must have its moments related to b_1 and b_2 in just this way. It follows then that since Table 1 was based on a 3-point distribution we can also use it to solve our present problem provided we enter it in an appropriate fashion. We may therefore summarize the derivation of the bounds of $\Pr(x \leqslant \alpha)$, when $\theta_3 (= \sqrt{\beta_1})$ and the range (b_1, b_2) of x are given, in the following way:

(i) Calculate both

$$\Delta_1' = -\frac{(b_1^2 - b_1\theta_3 - 1)(\alpha^2 - \alpha\theta_3 - 1)}{1 + \alpha b_1} \quad \text{and} \quad \Delta_2' = -\frac{(b_2^2 - b_2\theta_3 - 1)(\alpha^2 - \alpha\theta_3 - 1)}{1 + \alpha b_2}. \tag{11}$$

(ii) Observe which of these quantities lies between 0 and

$$-(b_1^2 - b_1\theta_3 - 1)(b_2^2 - b_2\theta_3 - 1)(1 + b_1 b_2)^{-1}.$$

(iii) Adopt this as a value of Δ to enter Table 1.

Whether any gain, apart from some insight into the solution, is obtained from this method of approach depends on whether we are led to a part of Table 1 where interpolation with respect to Δ is easy. (In this connexion it should be noted that interpolation between $\Delta = 0$ and 0·5 and in certain places between $\Delta = 16$ and ∞ is not to be recommended.)

If the range is limited at one end only, so that x is contained either in (a, ∞) or $(-\infty, a)$, say, the steps (i)–(iii) become

(i) Calculate $\Delta' = -\dfrac{(a^2 - a\theta_3 - 1)(\alpha^2 - \alpha\theta_3 - 1)}{1 + \alpha a}$.

(ii) If Δ' is positive, enter Table 1 with $\Delta = \Delta'$.

(iii) If Δ' is negative adopt the Tchebychef solution based on two moments (which is the same as entering Table 1 with $\Delta = \infty$; cf. equation (4), above).

We wish to acknowledge some helpful suggestions from the referees relating to the way in which the tables are presented.

REFERENCES

BERGE, P. O. (1932). Uber das Theorem von Tchebycheff und andere Grenzen einer Wahrscheinlichkeitsfunktion. *Skand. Actuar. Tidskr*, **15**, 65–77.

MALLOWS, C. L. (1956). Generalizations of Tchebycheff's inequalities. *J.R. Statist. Soc.* B, **18**, 139–68.

USPENSKY, J. V. (1937). *Introduction to Mathematical Probability*. New York and London: The McGraw-Hill Book Company.

ZELEN, M. (1954). Bounds on a distribution function that are functions of moments to order four. *J. Res. Nat. Bur. Stand.* **53**, 377–81.

Table 1. *Lower bound (L) and upper bound (U) of probability integral,* $\Pr(x \leqslant \alpha)$

(The variable, x, throughout is expressed in standard measure; its third moment is denoted by θ_3 or $\sqrt{\beta_1}$ and fourth moment by θ_4 or β_2; the quantity $\Delta = (\theta_4 - \theta_3^2 - 1) = (\beta_2 - \beta_1 - 1)$ is essentially positive.)

$\theta_3 = \sqrt{\beta_1} = 0$

$\theta_4 = \beta_2 \rightarrow$	1		1·5		2		3		4		5		7		9		13		17		∞	
$\Delta \rightarrow$	0		0·5		1		2		3		4		6		8		12		16		∞	
α	L	U	L	U	L	U	L	U	L	U	L	U	L	U	L	U	L	U	L	U	L	U
−10	0	0	0	0·000	0	0·000	0	0·000	0	0·000	0	0·000	0	0·001	0	0·001	0	0·001	0	0·001	0	0·010
− 9	0	0	0	·000	0	·000	0	·000	0	·000	0	·001	0	·001	0	·001	0	·002	0	·002	0	·012
− 8	0	0	0	·000	0	·000	0	·000	0	·001	0	·001	0	·001	0	·002	0	·003	0	·003	0	·015
− 7	0	0	0	·000	0	·000	0	·001	0	·001	0	·002	0	·002	0	·003	0	·004	0	·005	0	·020
− 6	0	0	0	·000	0	·001	0	·002	0	·002	0	·003	0	·004	0	·005	0	·007	0	·009	0	·027
− 5	0	0	0	·001	0	·002	0	·003	0	·005	0	·006	0	·008	0	·010	0	·014	0	·016	0	·038
− 4	0	0	0	·002	0	·004	0	·008	0	·011	0	·014	0	·018	0	·022	0	·028	0	·032	0	·059
− 3	0	0	0	·007	0	·014	0	·024	0	·032	0	·038	0	·048	0	·056	0	·065	0	·071	0	·100
− 2·5	0	0	0	0·016	0	0·029	0	0·048	0	0·061	0	0·071	0	0·084	0	0·094	0	0·105	0	0·111	0	0·138
− 2·0	0	0	0	·043	0	·071	0	·105	0	·125	0	·138	0	·154	0	·163	0	·174	0	·180	0	·200
− 1·5	0	0	0	·157	0	·208	0	·248	0	·265	0	·275	0	·285	0	·290	0	·296	0	·299	0	·308
− 1·0	—	—	0	·500	0	·500	0	·500	0	·500	0	·500	0	·500	0	·500	0	·500	0	·500	0	·500
− 0·5	0·500	0·500	0·172	·593	0·088	·640	0·036	·689	0·019	·715	0·012	·731	0·005	·750	0·003	·760	0·001	·772	0·001	·779	0	·800
0·0	0·500	0·500	0·333	0·667	0·250	0·750	0·167	0·833	0·125	0·875	0·100	0·900	0·071	0·929	0·056	0·944	0·038	0·962	0·029	0·971	0	1
0·5	0·500	0·500	0·407	0·824	0·360	0·912	0·311	0·964	0·285	0·981	0·269	0·988	0·250	0·995	0·240	0·997	0·228	0·999	0·221	0·999	0·200	1
1·0	—	—	·500	1	·500	1	·500	1	·500	1	·500	1	·500	1	·500	1	·500	1	·500	1	·500	1
1·5	1	1	·843	1	·792	1	·752	1	·735	1	·725	1	·715	1	·710	1	·704	1	·701	1	·692	1
2·0	1	1	·957	1	·929	1	·895	1	·875	1	·862	1	·846	1	·837	1	·826	1	·820	1	·800	1
2·5	1	1	·984	1	·971	1	·952	1	·939	1	·929	1	·916	1	·907	1	·895	1	·889	1	·862	1
3	1	1	0·993	1	0·986	1	0·976	1	0·968	1	0·962	1	0·952	1	0·944	1	0·935	1	0·929	1	0·900	1
4	1	1	·998	1	·996	1	·992	1	·989	1	·986	1	·982	1	·978	1	·972	1	·968	1	·941	1
5	1	1	·999	1	·998	1	·997	1	·995	1	·994	1	·992	1	·990	1	·986	1	·984	1	·962	1
6	1	1	1·000	1	·999	1	·998	1	·998	1	·997	1	·996	1	·995	1	·993	1	·991	1	·973	1
7	1	1	1·000	1	1·000	1	·999	1	·999	1	·998	1	·998	1	·997	1	·996	1	·995	1	·980	1
8	1	1	1·000	1	1·000	1	1·000	1	·999	1	·999	1	·999	1	·998	1	·997	1	·997	1	·985	1
9	1	1	1·000	1	1·000	1	1·000	1	1·000	1	·999	1	·999	1	·999	1	·998	1	·998	1	·988	1
10	1	1	1·000	1	1·000	1	1·000	1	1·000	1	1·000	1	·999	1	·999	1	·999	1	·999	1	·990	1

(For $\Delta = 0$, both L and U change from 0 to 0·500 at $\alpha = \lambda_1 = -1$ and from 0·500 to 1 at $\alpha = \lambda_2 = 1$.)

Table 1 (*cont*). *Lower bound* (L) *and upper bound* (U) *of probability integral*, $\Pr(x \leqslant \alpha)$

$\theta_3 = \sqrt{\beta_1} = 1$

$\theta_1 = \beta_2 \rightarrow$	2		2·5		3		4		5		6		8		10		14		18		∞	
$\Delta \rightarrow$	0		0·5		1		2		3		4		6		8		12		16		∞	
α	L	U	L	U	L	U	L	U	L	U	L	U	L	U	L	L	L	U	L	U	L	L
−10	0	0	0	0·000	0	0·000	0	0·000	0	0·000	0	0·000	0	0·000	0	0·001	0	0·001	0	0·001	0	0·010
− 9	0	0	0	·000	0	·000	0	·000	0	·000	0	·000	0	·001	0	·001	0	·001	0	·002	0	·012
− 8	0	0	0	·000	0	·000	0	·000	0	·001	0	·001	0	·001	0	·001	0	·002	0	·003	0	·015
− 7	0	0	0	·000	0	·000	0	·001	0	·001	0	·001	0	·002	0	·002	0	·003	0	·004	0	·020
− 6	0	0	0	·000	0	·001	0	·001	0	·002	0	·002	0	·003	0	·004	0	·006	0	·007	0	·027
− 5	0	0	0	·001	0	·001	0	·002	0	·003	0	·004	0	·006	0	·008	0	·010	0	·013	0	·038
− 4	0	0	0	·001	0	·003	0	·005	0	·007	0	·009	0	·013	0	·016	0	·021	0	·025	0	·059
− 3	0	0	0	·004	0	·008	0	·014	0	·020	0	·025	0	·033	0	·040	0	·050	0	·057	0	·100
2·5	0	0	0	0·009	0	0·015	0	0·027	0	0·037	0	0·045	0	0·058	0	0·068	0	0·082	0	0·091	0	0·138
− 2·0	0	0	0	·018	0	·033	0	·057	0	·075	0	·089	0	·109	0	·123	0	·141	0	·152	0	·200
− 1·5	0	0	0	·054	0	·092	0	·142	0	·173	0	·195	0	·222	0	·238	0	·258	0	·269	0	·308
− 1·0	0	0	0	·250	0	·333	0	·400	0	·429	0	·444	0	·462	0	·471	0	·480	0	·485	0	·500
− 0·5	0·724	0·724	0·031	·759	0·010	·772	0·003	·783	0·001	·788	0·001	·791	0·000	·794	0·000	·795	0·000	·797	0·000	·798	0	·800
0·0	0·724	0·724	0·459	0·793	0·333	0·833	0·213	0·880	0·155	0·905	0·122	0·922	0·085	0·942	0·065	0·954	0·044	0·967	0·033	0·974	0	1
0·5	0·724	0·724	0·603	0·832	0·530	0·886	0·444	0·937	0·395	0·960	0·363	0·972	0·323	0·985	0·299	0·990	0·271	0·995	0·255	0·997	0·200	1
1·0	·724	·724	·658	·908	·622	·955	·584	·984	·564	·992	·551	·996	·537	·998	·528	·999	·520	1·000	·515	1·000	·500	1
1·5	·724	·724	·702	·998	·698	1·000	·695	1·000	·694	1·000	·694	1·000	·693	1·000	·693	1·000	·693	1·000	·693	1·000	·692	1
2·0	1	1	·857	1	·833	1	·818	1	·813	1	·810	1	·806	1	·805	1	·803	1	·802	1	·800	1
2·5	1	1	·955	1	·932	1	·909	1	·898	1	·891	1	·883	1	·878	1	·873	1	·871	1	·862	1
3	1	1	0·983	1	0·971	1	0·956	1	0·945	1	0·938	1	0·929	1	0·924	1	0·917	1	0·914	1	0·900	1
4	1	1	·996	1	·993	1	·987	1	·983	1	·979	1	·973	1	·969	1	·963	1	·959	1	·941	1
5	1	1	·999	1	·997	1	·995	1	·993	1	·991	1	·988	1	·986	1	·982	1	·979	1	·962	1
6	1	1	·999	1	·999	1	·998	1	·997	1	·996	1	·994	1	·993	1	·991	1	·989	1	·973	1
7	1	1	1·000	1	·999	1	·999	1	·998	1	·998	1	·997	1	·996	1	·995	1	·994	1	·980	1
8	1	1	1·000	1	1·000	1	·999	1	·999	1	·999	1	·998	1	·998	1	·997	1	·996	1	·985	1
9	1	1	1·000	1	1·000	1	1·000	1	·999	1	·999	1	·999	1	·999	1	·998	1	·997	1	·988	1
10	1	1	1·000	1	1·000	1	1·000	1	1·000	1	1·000	1	·999	1	·999	1	·999	1	·998	1	·990	1

(For $\Delta = 0$, both L and U change from 0 to 0·724 at $\alpha = \lambda_1 = -0{\cdot}618$ and from 0·724 to 1 at $\alpha = \lambda_2 = 1{\cdot}618$.)

Table 1 (*cont.*)

$\theta_2 = \sqrt{\beta_1} = 2$

$\theta_1 = \beta_2 \rightarrow$	5		5·5		6		7		8		9		11		13		17		21		∞	
$\Delta \rightarrow$	0		0·5		1		2		3		4		6		8		12		16		∞	
α	L	U	L	U	L	U	L	U	L	U	L	U	L	U	L	U	L	U	L	U	L	U
−10	0	0	0	0·000	0	0·000	0	0·000	0	0·000	0	0·000	0	0·000	0	0·001	0	0·001	0	0·001	0	0·010
− 9	0	0	0	·000	0	·000	0	·000	0	·000	0	0·000	0	·001	0	·001	0	·001	0	·001	0	·012
− 8	0	0	0	·000	0	·000	0	·000	0	·000	0	·001	0	·001	0	·001	0	·002	0	·002	0	·015
− 7	0	0	0	·000	0	·000	0	·001	0	·001	0	·001	0	·001	0	·002	0	·003	0	·003	0	·020
− 6	0	0	0	·000	0	·000	0	·001	0	·001	0	·002	0	·002	0	·003	0	·005	0	·006	0	·027
− 5	0	0	0	·000	0	·001	0	·002	0	·002	0	·003	0	·005	0	·006	0	·008	0	·010	0	·038
− 4	0	0	0	·001	0	·002	0	·004	0	·005	0	·007	0	·010	0	·012	0	·016	0	·020	0	·059
− 3	0	0	0	·002	0	·005	0	·009	0	·013	0	·017	0	·023	0	·029	0	·038	0	·045	0	·100
− 2·5	0	0	0	0·005	0	0·009	0	0·017	0	0·024	0	0·030	0	0·040	0	0·049	0	0·062	0	0·072	0	0·138
− 2·0	0	0	0	·010	0	·019	0	·034	0	·047	0	·058	0	·076	0	·090	0	·110	0	·124	0	·200
− 1·5	0	0	0	·025	0	·047	0	·081	0	·108	0	·129	0	·160	0	·182	0	·210	0	·228	0	·308
− 1·0	0	0	0	·100	0	·167	0	·250	0	·300	0	·333	0	·375	0	·400	0	·429	0	·444	0	·500
− 0·5	0	0	0	·727	0	·762	0	·780	0	·787	0	·790	0	·793	0	·795	0	·797	0	·798	0	·800
0·0	0·854	0·854	0·544	0·877	0·394	0·894	0·250	0·917	0·181	0·931	0·141	0·941	0·097	0·954	0·073	0·962	0·049	0·972	0·036	0·978	0	1
0·5	0·854	0·854	0·751	0·886	0·677	0·909	0·578	0·937	0·514	0·954	0·469	0·965	0·409	0·978	0·372	0·985	0·327	0·991	0·301	0·995	0·200	1
1·0	·854	·854	·801	·901	·762	·929	·708	·958	·673	·973	·648	·982	·615	·990	·594	·994	·569	·997	·554	·999	·500	1
1·5	·854	·854	·821	·928	·800	·958	·772	·981	·756	·990	·745	·994	·731	·997	·723	·999	·714	·999	·709	1·000	·692	1
2·0	·854	·854	·834	·977	·825	·992	·816	·998	·812	·999	·809	1·000	·806	1·000	·805	1·000	·803	1·000	·802	1·000	·800	1
2·5	1	1	·864	1	·863	1	·863	1	·862	1	·862	1	·862	1	·862	1	·862	1	·862	1	·862	1
3	1	1	0·944	1	0·929	1	0·917	1	0·912	1	0·909	1	0·906	1	0·905	1	0·903	1	0·902	1	0·900	1
4	1	1	·991	1	·985	1	·976	1	·970	1	·966	1	·960	1	·957	1	·953	1	·950	1	·941	1
5	1	1	·998	1	·996	1	·992	1	·989	1	·987	1	·983	1	·980	1	·976	1	·974	1	·962	1
6	1	1	·999	1	·998	1	·997	1	·995	1	·994	1	·992	1	·990	1	·988	1	·986	1	·973	1
7	1	1	1·000	1	·999	1	·998	1	·998	1	·997	1	·996	1	·995	1	·993	1	·992	1	·980	1
8	1	1	1·000	1	1·000	1	·999	1	·999	1	·998	1	·998	1	·997	1	·996	1	·995	1	·985	1
9	1	1	1·000	1	1·000	1	1·000	1	·999	1	·999	1	·999	1	·998	1	·998	1	·997	1	·988	1
10	1	1	1·000	1	1·000	1	1·000	1	1·000	1	·999	1	·999	1	·999	1	·998	1	·998	1	·990	1

(For $\Delta = 0$, both the L and U change from 0 to 0·854 at $\alpha = \lambda_1 = -0{\cdot}414$ and from 0·854 to 1 at $\alpha = \lambda_2 = 2{\cdot}414$.)

Table 1 (*cont.*) *Lower bound (L) and upper bound (U) of probability integral,* $\Pr(x \leqslant \alpha)$

$\theta_3 = \sqrt{\beta_1} = 3$

$\theta_4 = \beta_2 \rightarrow$	10		10·5		11		12		13		14		16		18		22		26		∞	
$\Delta \rightarrow$	0		0·5		1		2		3		4		6		8		12		16		∞	
α	L	U	L	U	L	U	L	U	L	U	L	U	L	U	L	U	L	U	L	U	L	U
−10	0	0	0	0·000	0	0·000	0	0·000	0	0·000	0	0·000	0	0·000	0	0·000	0	0·001	0	0·001	0	0·010
− 9	0	0	0	·000	0	·000	0	·000	0	·000	0	·000	0	·001	0	·001	0	·001	0	·001	0	·012
− 8	0	0	0	·000	0	·000	0	·000	0	·000	0	·001	0	·001	0	·001	0	·001	0	·002	0	·015
− 7	0	0	0	·000	0	·000	0	·000	0	·001	0	·001	0	·001	0	·002	0	·002	0	·003	0	·020
− 6	0	0	0	·000	0	·000	0	·001	0	·001	0	·001	0	·002	0	·003	0	·004	0	·005	0	·027
− 5	0	0	0	·000	0	·001	0	·001	0	·002	0	·002	0	·004	0	·005	0	·007	0	·008	0	·038
− 4	0	0	0	·001	0	·001	0	·003	0	·004	0	·005	0	·007	0	·009	0	·013	0	·016	0	·059
− 3	0	0	0	·002	0	·003	0	·006	0	·009	0	·012	0	·017	0	·022	0	·029	0	·036	0	·100
− 2·5	0	0	0	0·003	0	0·006	0	0·011	0	0·016	0	0·021	0	0·029	0	0·036	0	0·048	0	0·057	0	0·138
− 2·0	0	0	0	·006	0	·012	0	·022	0	·031	0	·040	0	·054	0	·066	0	·085	0	·099	0	·200
− 1·5	0	0	0	·014	0	·028	0	·051	0	·070	0	·087	0	·114	0	·135	0	·167	0	·188	0	·308
− 1·0	0	0	0	·050	0	·091	0	·154	0	·200	0	·235	0	·286	0	·320	0	·364	0	·390	0	·500
− 0·5	0	0	0	·421	0	·552	0	·653	0	·696	0	·719	0	·744	0	·757	0	·771	0	·778	0	·800
0·0	0·916	0·916	0·592	0·925	0·432	0·932	0·276	0·942	0·200	0·950	0·156	0·956	0·107	0·964	0·080	0·969	0·053	0·976	0·039	0·981	0	1
0·5	0·916	0·916	0·839	0·927	0·778	0·936	0·685	0·950	0·619	0·959	0·569	0·966	0·498	0·976	0·451	0·982	0·391	0·989	0·354	0·992	0·200	1
1·0	·916	·916	·880	·930	·851	·942	·803	·957	·767	·967	·739	·975	·698	·983	·668	·988	·630	·994	·606	·996	·500	1
1·5	·916	·916	·895	·936	·877	·949	·849	·967	·829	·977	·813	·983	·790	·990	·775	·994	·755	·997	·743	·998	·692	1
2·0	·916	·916	·902	·945	·890	·962	·873	·979	·862	·987	·853	·991	·842	·996	·834	·997	·825	·999	·820	·999	·800	1
2·5	·916	·916	·906	·963	·899	·980	·889	·992	·884	·996	·880	·997	·875	·999	·873	·999	·869	1·000	·868	1·000	·862	1
3	0·916	0·916	0·910	0·993	0·907	0·998	0·904	0·999	0·903	1·000	0·902	1·000	0·902	1·000	0·901	1·000	0·901	1·000	0·901	1·000	0·900	1
4	1	1	·971	1	·962	1	·953	1	·950	1	·948	1	·946	1	·945	1	·944	1	·943	1	·941	1
5	1	1	·995	1	·991	1	·985	1	·981	1	·978	1	·975	1	·972	1	·969	1	·968	1	·962	1
6	1	1	·998	1	·997	1	·994	1	·993	1	·991	1	·988	1	·986	1	·984	1	·982	1	·973	1
7	1	1	·999	1	·999	1	·998	1	·997	1	·996	1	·994	1	·993	1	·991	1	·990	1	·980	1
8	1	1	1·000	1	·999	1	·999	1	·998	1	·998	1	·997	1	·996	1	·995	1	·994	1	·985	1
9	1	1	1·000	1	1·000	1	·999	1	·999	1	·999	1	·998	1	·998	1	·997	1	·996	1	·988	1
10	1	1	1·000	1	1·000	1	1·000	1	·999	1	·999	1	·999	1	·999	1	·998	1	·997	1	·990	1

(For $\Delta = 0$, both L and U change from 0 to 0·916 at $\alpha = \lambda_1 = -0{\cdot}303$ and from 0·916 to 1 at $\alpha = \lambda_2 = 3{\cdot}303$.)

Table 1 (*cont.*)

$\theta_2 = \sqrt{\beta_1} = 4$

$\theta_1 = \beta_2 \rightarrow$	17		17·5		18		19		20		21		23		25		29		33		∞	
$\Delta \rightarrow$	0		0·5		1		2		3		4		6		8		12		16		∞	
α	L	U	L	U	L	U	L	U	L	U	L	U	L	U	L	U	L	U	L	U	L	U
−10	0	0	0	0·000	0	0·000	0	0·000	0	0·000	0	0·000	0	0·000	0	0·000	0	0·001	0	0·001	0	0·010
− 9	0	0	0	·000	0	·000	0	·000	0	·000	0	·000	0	·000	0	·001	0	·001	0	·001	0	·012
− 8	0	0	0	·000	0	·000	0	·000	0	·000	0	·000	0	·001	0	·001	0	·001	0	·002	0	·015
− 7	0	0	0	·000	0	·000	0	·000	0	·001	0	·001	0	·001	0	·001	0	·002	0	·002	0	·020
− 6	0	0	0	·000	0	·000	0	·001	0	·001	0	·001	0	·002	0	·002	0	·003	0	·004	0	·027
− 5	0	0	0	·000	0	·001	0	·001	0	·001	0	·002	0	·003	0	·004	0	·005	0	·007	0	·038
− 4	0	0	0	·001	0	·001	0	·002	0	·003	0	·004	0	·006	0	·007	0	·010	0	·013	0	·059
− 3	0	0	0	·001	0	·002	0	·005	0	·007	0	·009	0	·013	0	·017	0	·023	0	·029	0	·100
− 2·5	0	0	0	0·002	0	0·004	0	0·008	0	0·012	0	0·015	0	0·022	0	0·028	0	0·038	0	0·046	0	0·138
− 2·0	0	0	0	·004	0	·008	0	·015	0	·022	0	·028	0	·040	0	·050	0	·066	0	·080	0	·200
− 1·5	0	0	0	·009	0	·018	0	·034	0	·048	0	·061	0	·083	0	·102	0	·131	0	·153	0	·308
− 1·0	0	0	0	·029	0	·056	0	·100	0	·136	0	·167	0	·214	0	·250	0	·300	0	·333	0	·500
− 0·5	0	0	0	·229	0	·356	0	·492	0	·565	0	·610	0	·662	0	·692	0	·725	0	·742	0	·800
0·0	0·947	0·947	0·618	0·951	0·454	0·954	0·293	0·959	0·213	0·963	0·167	0·967	0·115	0·972	0·086	0·975	0·057	0·980	0·042	0·983	0	1
0·5	0·947	0·947	0·891	0·952	0·842	0·956	0·763	0·962	0·702	0·967	0·653	0·971	0·579	0·978	0·526	0·982	0·456	0·988	0·411	0·991	0·200	1
1·0	·947	·947	·923	·953	·902	·957	·865	·965	·834	·971	·808	·975	·767	·982	·736	·986	·691	·991	·661	·994	·500	1
1·5	·947	·947	·933	·954	·921	·960	·899	·968	·882	·974	·867	·979	·843	·985	·825	·989	·799	·994	·782	·996	·692	1
2·0	·947	·947	·938	·956	·930	·963	·916	·973	·904	·979	·895	·984	·881	·990	·870	·993	·855	·996	·845	·998	·800	1
2·5	·947	·947	·940	·959	·934	·968	·925	·979	·918	·985	·912	·989	·903	·994	·897	·996	·889	·998	·883	·999	·862	1
3	0·947	0·947	0·942	0·966	0·938	0·976	0·931	0·987	0·927	0·992	0·923	0·995	0·918	0·997	0·915	0·998	0·911	0·999	0·909	1·000	0·900	1
4	·947	·947	·945	·997	·944	·999	·943	1·000	·942	1·000	·942	1·000	·942	1·000	·942	1·000	·941	1·000	·941	1·000	·941	1
5	1	1	·983	1	·976	1	·971	1	·968	1	·967	1	·965	1	·964	1	·963	1	·963	1	·962	1
6	1	1	·996	1	·994	1	·990	1	·987	1	·985	1	·983	1	·981	1	·979	1	·978	1	·973	1
7	1	1	·999	1	·998	1	·996	1	·995	1	·993	1	·991	1	·990	1	·988	1	·987	1	·980	1
8	1	1	1·000	1	·999	1	·998	1	·997	1	·997	1	·996	1	·995	1	·993	1	·992	1	·985	1
9	1	1	1·000	1	1·000	1	·999	1	·999	1	·998	1	·998	1	·997	1	·996	1	·995	1	·988	1
10	1	1	1·000	1	1·000	1	·999	1	·999	1	·999	1	·999	1	·998	1	·997	1	·997	1	·990	1

(For $\Delta = 0$, both L and U change from 0 to 0·947 at $\alpha = \lambda_1 = -0{\cdot}236$ and from 0·947 to 1 at $\alpha = \lambda_2 = 4{\cdot}236$.)

Table 1 (*cont.*) *Lower bound* (L) *and upper bound* (U) *of probability integral,* $\Pr(x \leqslant \alpha)$

$\theta_2 = \sqrt{\beta_1} = 8$

$\theta_1 = \beta_2 \rightarrow$	65		65·5		66		67		68		69		71		73		77		81		∞	
$\Delta \rightarrow$	0		0·5		1		2		3		4		6		8		12		16		∞	
α	L	U	L	U	L	U	L	U	L	U	L	U	L	U	L	U	L	U	L	U	L	U
−10	0	0	0	0·000	0	0·000	0	0·000	0	0·000	0	0·000	0	0·000	0	0·000	0	0·000	0	0·000	0	0·010
− 9	0	0	0	·000	0	·000	0	·000	0	·000	0	·000	0	·000	0	·000	0	·001	0	·001	0	·012
− 8	0	0	0	·000	0	·000	0	·000	0	·000	0	·000	0	·000	0	·000	0	·001	0	·001	0	·015
− 7	0	0	0	·000	0	·000	0	·000	0	·000	0	·000	0	·001	0	·001	0	·001	0	·001	0	·020
− 6	0	0	0	·000	0	·000	0	·000	0	·000	0	·001	0	·001	0	·001	0	·002	0	·002	0	·027
− 5	0	0	0	·000	0	·000	0	·000	0	·001	0	·001	0	·001	0	·002	0	·003	0	·004	0	·038
− 4	0	0	0	·000	0	·000	0	·001	0	·001	0	·002	0	·003	0	·003	0	·005	0	·006	0	·059
− 3	0	0	0	·000	0	·001	0	·002	0	·003	0	·004	0	·006	0	·007	0	·010	0	·014	0	·100
− 2·5	0	0	0	0·001	0	0·002	0	0·003	0	0·005	0	0·006	0	0·009	0	0·012	0	0·017	0	0·021	0	0·138
− 2·0	0	0	0	·001	0	·003	0	·005	0	·008	0	·011	0	·015	0	·020	0	·029	0	·036	0	·200
− 1·5	0	0	0	·003	0	·006	0	·011	0	·016	0	·021	0	·031	0	·040	0	·056	0	·070	0	·308
− 1·0	0	0	0	·008	0	·015	0	·029	0	·043	0	·056	0	·079	0	·100	0	·136	0	·167	0	·500
− 0·5	0	0	0	·045	0	·085	0	·153	0	·210	0	·257	0	·332	0	·389	0	·469	0	·524	0	·800
0·0	0·985	0·985	0·652	0·985	0·486	0·986	0·320	·986	0·237	0·987	0·187	0·987	0·131	0·988	0·100	0·989	0·067	0·990	0·050	0·991	0	1
0·5	0·985	0·985	0·964	0·985	0·944	0·986	0·907	0·986	0·873	0·987	0·842	0·988	0·789	0·989	0·744	0·990	0·672	0·991	0·616	0·992	0·200	1
1·0	·985	·985	·978	·985	·971	·986	·957	·987	·944	·987	·932	·988	·910	·989	·890	·990	·855	·992	·826	·993	·500	1
1·5	·985	·985	·981	·986	·977	·986	·970	·987	·963	·987	·957	·988	·945	·989	·934	·990	·914	·992	·898	·993	·692	1
2·0	·985	·985	·983	·986	·980	·986	·976	·987	·971	·988	·967	·988	·959	·990	·952	·991	·940	·993	·930	·994	·800	1
2·5	·985	·985	·983	·986	·982	986	·978	·987	·975	·988	·972	·989	·967	·990	·962	·991	·954	·993	·946	·994	·862	1
3	0·985	0·985	0·984	0·986	0·983	0·986	0·980	0·987	0·978	0·988	0·976	0·989	0·972	0·991	0·968	0·992	0·962	0·994	0·957	0·995	0·900	1
4	·985	·985	·984	·986	·983	·987	·982	·988	·980	·989	·979	·990	·977	·992	·974	·993	·971	·995	·968	·996	·941	1
5	·985	·985	·984	·986	·984	·987	·983	·989	·982	·991	·981	·992	·979	·994	·978	·995	·976	·997	·974	·998	·962	1
6	·985	·985	·985	·987	·984	·989	·983	·992	·983	·993	·982	·995	·981	·996	·980	·997	·979	·999	·978	·999	·973	1
7	·985	·985	·985	·990	·984	·993	·984	·996	·983	·997	·983	·998	·983	·999	·982	·999	·982	1·000	·981	1·000	·980	1
8	·985	·985	·985	1·000	·985	1·000	·985	1·000	·985	1·000	·985	1·000	·985	1·000	·985	1·000	·985	1·000	·985	1·000	·985	1
9	1	1	·995	1	·993	1	·991	1	·990	1	·990	1	·989	1	·989	1	·989	1	·988	1	·988	1
10	1	1	·999	1	·998	1	·996	1	·995	1	·995	1	·994	1	·993	1	·992	1	·992	1	·990	1

(For $\Delta = 0$, both L and U change from 0 to 0·985 at $\alpha = \lambda_1 = -0{\cdot}123$ and from 0·985 to 1 at $\alpha = \lambda_2 = 8{\cdot}123$.)

Biometrika (1960), **47**, 3 *and* 4, *p.* 411
Printed in Great Britain

Normal approximation to the chi-square and non-central *F* probability functions

BY NORMAN C. SEVERO AND MARVIN ZELEN*
University of Buffalo and National Bureau of Standards

1. INTRODUCTION

Although extensive tabulations of many probability functions are now available, there continues to be interest in developing easier and more accurate approximation formulae. These approximations are especially useful to the occasional computer who does not have ready access to extensive tables. In this paper we present normal approximations to the χ^2 and non-central F probability functions. The χ^2 approximation is an improved Wilson–Hilferty normalized deviate. The non-central F approximation employs a cube-root transformation and a 'Geary type' solution.

2. THE χ^2 APPROXIMATION

The χ^2 probability function with ν degrees of freedom is defined by

$$P(\chi^2|\nu) = [2^{\frac{1}{2}\nu}\Gamma(\tfrac{1}{2}\nu)]^{-1}\int_0^{\chi^2} t^{\frac{1}{2}\nu-1}e^{-\frac{1}{2}t}\,dt$$

for $0 \leqslant \chi^2 < \infty$. Two well-known approximations to this probability function are due to R. A. Fisher and to E. B. Wilson and M. M. Hilferty (see Merrington, 1941). Fisher's approximation takes $\sqrt{(2\chi^2)}$ to be normally distributed about $\sqrt{(2\nu-1)}$ with unit standard deviation. The Wilson–Hilferty approximation is based on the fact that $(\chi^2/\nu)^{\frac{1}{3}}$ is approximately normally distributed with mean $1-2/(9\nu)$ and standard deviation $\sqrt{\{2/(9\nu)\}}$.

Let the standardized normal probability function be denoted by $P(x)$; i.e.

$$P(x) = \int_{-\infty}^{x}\frac{1}{\sqrt{(2\pi)}}e^{-\frac{1}{2}t^2}\,dt.$$

Then these two well-known approximations to the $100p$ percentage point of the χ^2-distribution may be written

$$\chi^2_p(\nu) \doteq \tfrac{1}{2}\{x_p+\sqrt{(2\nu-1)}\}^2 \quad \text{(Fisher)} \tag{1}$$

and

$$\chi^2_p(\nu) \doteq \nu\left\{1-\frac{2}{9\nu}+x_p\sqrt{\frac{2}{9\nu}}\right\}^3 \quad \text{(Wilson–Hilferty)}, \tag{2}$$

where x_p is such that

$$P(x_p) = 1-p. \tag{3}$$

Merrington (1941) presented a numerical comparison of approximations (1) and (2) with the exact values for $\nu \geqslant 30$ (cf. Table 8 of Pearson & Hartley, 1954). We have reproduced Merrington's comparison table in our Table 1, which also includes the values obtained by the new approximation given below.

Let the probability function for

$$X = \frac{(\chi^2/\nu)^{\frac{1}{3}}-(1-2/9\nu)}{(2/9\nu)^{\frac{1}{2}}} \tag{4}$$

* Now at the University of Maryland.

Table 1. *Comparative table of percentage points of* χ^2

ν \ P	0·995	0·990	0·950	0·900	0·750	0·500	0·250	0·100	0·050	0·010	0·005
30 (*a*)	13·032	14·337	18·218	20·477	24·547	29·500	34·908	40·165	43·487	50·075	52·603
(*b*)	13·744	14·925	18·491	20·604	24·486	29·338	34·793	40·246	43·767	50·914	53·713
(*c*)	13·7867	14·9535	18·4926	20·5992	24·4776	29·3360	34·7998	40·2560	43·7729	50·8922	53·6720
(*d*)	13·779	14·949	18·493	20·601	24·478	29·336	34·799	40·256	43·772	50·888	53·665
40 (*a*)	19·923	21·529	26·233	28·931	33·732	39·500	45·722	51·712	55·473	62·883	65·712
(*b*)	20·669	22·139	26·508	29·055	33·668	39·337	45·610	51·796	55·753	63·710	66·802
(*c*)	20·7065	22·1643	26·5093	29·0505	33·6603	39·3354	45·6160	51·8050	55·7585	63·6907	66·7659
(*d*)	20·702	22·161	26·510	29·051	33·660	39·334	45·615	51·804	55·757	63·688	66·763
50 (*a*)	27·188	29·059	34·487	37·570	43·016	49·500	56·439	63·072	67·219	75·353	78·447
(*b*)	27·957	29·685	34·763	37·693	42·949	49·336	56·382	63·159	67·501	76·172	79·523
(*c*)	27·9907	29·7067	34·7642	37·6886	42·9421	49·3349	56·3336	63·1671	67·5048	76·1539	79·4900
(*d*)	27·988	29·706	34·766	37·691	42·943	49·336	56·334	63·168	67·506	76·155	79·488
75 (*a*)	46·375	48·809	55·775	59·678	66·494	74·500	82·961	90·965	95·931	105·603	109·259
(*b*)	47·178	49·457	56·054	59·799	66·422	74·335	82·853	91·055	96·214	106·408	110·313
(*c*)	47·2059	49·4748	56·0540	59·7944	66·4167	74·3343	82·8582	91·062	96·217	106·393	110·286
(*d*)	47·206	49·476	56·057	59·797	66·418	74·334	82·858	91·061	96·218	106·394	110·288
100 (*a*)	66·481	69·389	77·649	82·243	90·213	99·500	109·242	118·400	124·056	135·023	139·154
(*b*)	67·303	70·049	77·929	82·362	90·138	99·335	109·137	118·493	124·340	135·820	140·193
(*c*)	67·3276	70·0648	77·9295	82·3581	90·1332	99·3341	109·141	118·498	124·342	135·807	140·169
(*d*)	67·328	70·068	77·932	82·361	90·134	99·335	109·142	118·499	124·344	135·808	140·173
150 (*a*)	108·277	111·980	122·411	128·161	138·064	149·500	161·390	172·481	179·295	192·432	197·358
(*b*)	109·122	112·655	122·692	128·278	137·987	149·336	161·288	172·577	179·579	193·219	198·380
(*c*)	—	112·668	—	128·275	137·983	149·334	161·291	172·581	—	193·208	—
(*d*)	109·146	112·671	122·692	128·277	137·984	149·336	161·292	172·584	179·580	193·209	198·363
200 (*a*)	151·365	155·737	167·997	174·722	186·255	199·500	213·200	225·920	233·709	248·675	254·270
(*b*)	152·224	156·421	168·279	174·838	186·175	199·334	213·099	226·017	233·993	249·455	255·305
(*c*)	—	156·432*	—	174·835*	186·172*	199·334*	213·102*	226·021*	—	249·446*	—
(*d*)	152·244	156·434	168·278	174·838	186·172	199·334	213·104	226·020	233·994	249·446	255·268
X_p	−2·57583	−2·32635	−1·64485	−1·28155	−0·67449	0·0000	0·67449	1·28155	1·64485	2·32635	2·57583

(*a*) From R. A. Fisher's formula. (*b*) From E. B. Wilson and M. M. Hilferty's formula. (*c*) From Miss Thompson's table with the exception of * values which were taken from Campbell (1923). (*d*) New approximation, equation (13).

be denoted by $P_1(x|\nu)$. Then using the terms of the Gram–Charlier A series to $O(\nu^{-\frac{1}{2}})$

$$P_1(x|\nu) = P(x) - \frac{\gamma_1 P^{(3)}(x)}{3!} + \frac{\gamma_2 P^{(4)}(x)}{4!}, \tag{5}$$

where $P^{(n)}(x)$ is the nth derivative of $P(x)$ and γ_1 and γ_2 are the measures of skewness and excess for X. These moments are given approximately by (Abdel-Aty, 1954)

$$\gamma_1 = \frac{2^{\frac{3}{2}}}{3^3 \nu^{\frac{1}{2}}}, \quad \gamma_2 = -\frac{4}{9\nu}. \tag{6}$$

Consider the approximation to the normal probability function $P(x+h)$ as given by the first two terms of the Taylor expansion about x; i.e.

$$P(x+h) = P(x) + hP^{(1)}(x). \tag{7}$$

Our new approximation involves finding that value of h, say $h_\nu(x)$, which for given x and ν gives $P_1(x|\nu) = P(x+h_\nu(x))$. This is found approximately by equating the right-hand sides of (5) and (7). In this way we obtain

$$\begin{aligned} h_\nu(x) &\doteq \frac{1}{P^{(1)}(x)} \left\{ -\frac{\gamma_1 P^{(3)}(x)}{3!} + \frac{\gamma_2 P^{(4)}(x)}{4!} \right\} \\ &= -\frac{2}{3^3 \nu} \left\{ \frac{2^{\frac{3}{2}}(x^2-1)}{3\sqrt{\nu}} - \frac{(x^3-3x)}{4} \right\}. \end{aligned} \tag{8}$$

We shall drop the argument x and write simply h_ν for $h_\nu(x)$. Thus our approximation to the probability function of χ^2 becomes

$$P(\chi^2|\nu) \doteq P(x+h_\nu), \tag{9}$$

where x and h_ν are given by (4) and (8), respectively. For the $100p$ percentage point we obtain the approximation

$$\chi_p^2(\nu) \doteq \nu \left\{ 1 - \frac{2}{9\nu} + (x_p - h_\nu)\sqrt{\frac{2}{9\nu}} \right\}^3, \tag{10}$$

where x_p is given by (3) and h_ν is the value of (8) at the point $x = x_p$.

A good approximation to equation (8) in the neighbourhood of $\nu = \nu_0$ may be obtained by noticing that

$$h_\nu = \frac{\nu_0}{\nu} h_{\nu_0} + R, \tag{11}$$

where

$$R = \frac{2^{\frac{5}{2}}(x^2-1)}{3^4 \nu} \left[\frac{1}{\sqrt{\nu_0}} - \frac{1}{\sqrt{\nu}} \right]. \tag{12}$$

Therefore in the neighbourhood of $\nu_0 = 60$ we may use

$$h_\nu \doteq \frac{60}{\nu} h_{60}, \tag{13}$$

where h_{60} is given in Table 2.

Table 2. *Factors for use in equation* (13)

x	h_{60}	x	h_{60}	x	h_{60}
−3·5	−0·0118	−1·0	0·0006	1·5	−0·0005
−3·0	− ·0067	−0·5	·0006	2·0	·0002
−2·5	− ·0033	0·0	·0002	2·5	·0017
−2·0	− ·0010	0·5	− ·0003	3·0	·0043
−1·5	·0001	1·0	− ·0006	3·5	·0082

Equation (13) with linear interpolation in Table 2 was used to obtain the new approximation values for the percentage points in Table 1. Notice that $R = 0$ whenever $\nu = \nu_0$ or $x = \pm 1$ so that in Table 1 we would expect the best results when $\nu = 60$ or $x = \pm 1$.

In Table 3 we show some comparisons for selected small values of ν. The (b) values in this table were obtained by using the approximation

$$h_\nu = \frac{20}{\nu} h_{20}, \tag{14}$$

where the appropriate h_{20} values were linearly interpolated from Table 4.

Table 3. *Comparative table of percentage points of χ^2 for small values of ν*

ν \ P	0·995	0·950	0·750	0·500	0·250	0·050	0·005
5 (a)	0·412	1·145	2·675	4·351	6·626	11·070	16·750
(b)	0·381	1·134	2·675	4·354	6·631	11·076	16·700
10 (a)	2·156	3·940	6·737	9·342	12·549	18·307	25·188
(b)	2·136	3·937	6·738	9·342	12·550	18·313	25·178
20 (a)	7·434	10·851	15·452	19·337	23·828	31·410	39·997
(b)	7·427	10·852	15·452	19·336	23·827	31·415	40·002

(a) Exact values. (b) New approximation, equation (14).

Table 4. *Factors for use in equation* (14)

x	h_{20}	x	h_{20}	x	h_{20}
−3·5	−0·03876	−1·0	0·00185	1·5	−0·00202
−3·0	− ·02291	−0·5	·00186	2·0	− ·00049
−2·5	− ·01162	0·0	·00078	2·5	·00342
−2·0	− ·00419	0·5	− ·00069	3·0	·01042
−1·5	·00007	1·0	− ·00185	3·5	·02119

3. THE NON-CENTRAL F APPROXIMATION

The non-central F statistic is defined by

$$F' = \frac{\chi_1'^2/\nu_1}{\chi_2^2/\nu_2},$$

where $\chi_1'^2$ follows a non-central chi-square distribution with degrees of freedom ν_1 and non-centrality parameter λ; χ_2^2 follows a central chi-square distribution with degrees of freedom ν_2; and $\chi_1'^2$ and χ_2^2 are statistically independent. Patnaik (1949) has obtained an approximation to the non-central F distribution by fitting an F-distribution with the exact first two moments. In Table 5 we reproduce Patnaik's comparative table that showed values obtained by use of his approximation and exact values computed by Tang. We have also included in Table 5 values obtained by the approximation developed below (equation (17)). Patnaik pointed out that two-way interpolation (linear in one direction and four-point Lagrangian in the other) in the *Tables of the Incomplete Beta-Function* (K. Pearson, 1934) was necessary in order to obtain his approximate values. The values computed by our method involved only linear interpolation in the Normal tables.

Denote the probability integral of F' by $P(F'|\nu_1, \nu_2, \lambda)$. Let

$$Y_1' = [\chi_1'^2/(\nu_1+\lambda)]^{\frac{1}{3}}, \quad Y_2 = [\chi_2^2/\nu_2]^{\frac{1}{3}}, \quad c = \left(\frac{\nu_1+\lambda}{\nu_1}\right)^{\frac{1}{3}}. \tag{15}$$

Then the probability integral of F' becomes

$$P(F'|\nu_1, \nu_2, \lambda) = \Pr\left\{\frac{\chi_1'^2/\nu_1}{\chi_2^2/\nu_2} \leqslant F'\right\} = \Pr\{L \leqslant 0\},$$

Table 5. *Approximate and exact values of the F'-probability function,* $P(F'|\nu_1, \nu_2, \lambda)$

ν_1	ν_2	λ	F	Patnaik approximation	Exact	New approximation
3	10	4	3·708	0·752	0·745	0·750
		4	6·552	·919	·918	·919
		16	3·708	·203	·206	·202
		16	6·552	·520	·517	·520
3	20	4	3·098	0·706	0·700	0·707
		4	4·938	·889	·887	·889
		16	3·098	·119	·126	·119
		16	4·938	·350	·347	·349
5	10	6	3·326	0·731	0·731	0·734
		6	5·636	·913	·914	·914
		24	3·326	·157	·158	·155
		24	5·636	·463	·461	·463
5	20	6	2·711	0·665	0·664	0·669
		6	4·103	·869	·870	·871
		24	2·711	·064	·069	·064
		24	4·103	·244	·245	·244
8	10	9	3·072	0·715	0·714	0·716
		9	5·057	·909	·908	·909
		36	3·072	·117	·119	·117
		36	5·057	·409	·408	·409
8	30	9	2·266	0·581	0·578	0·581
		9	3·173	·815	·813	·815
		36	2·266	·014	·017	·014
		36	3·173	·085	·088	·085

where $L = cY_1' - (F')^{\frac{1}{3}}Y_2$. The mean and variance of Y_1' are given approximately by Abdel-Aty (1954), as

$$1 - \frac{2}{9}\left[\frac{\nu_1+2\lambda}{(\nu_1+\lambda)^2}\right] \quad \text{and} \quad \frac{2}{9}\left[\frac{\nu_1+2\lambda}{(\nu_1+\lambda)^2}\right],$$

respectively. Because the distributions of Y_1' and Y_2 are each approximately normal, it follows that L is approximately normal with mean and variance given by

$$\mu_L = c - \frac{2c}{9}\left[\frac{\nu_1+2\lambda}{(\nu_1+\lambda)^2}\right] - (F')^{\frac{1}{3}}\left[1 - \frac{2}{9\nu_2}\right]$$

and

$$\sigma_L^2 = \frac{2c^2(\nu_1+2\lambda)}{9(\nu_1+\lambda)^2} + \frac{2(F')^{\frac{2}{3}}}{9\nu_2},$$

respectively. Therefore

$$\Pr\{L \leqslant 0\} \doteq P(-\mu_L/\sigma_L)$$

and so the normal approximation to the non-central F distribution becomes

$$P(F'|\nu_1, \nu_2, \lambda) \doteq P(x), \tag{16}$$

where

$$x = \frac{\left(\frac{\nu_1 F'}{\nu_1+\lambda}\right)^{\frac{1}{3}}\left(1-\frac{2}{9\nu_2}\right) - \left[1-\frac{2(\nu_1+2\lambda)}{9(\nu_1+\lambda)^2}\right]}{\left[\frac{2(\nu_1+2\lambda)}{9(\nu_1+\lambda)^2}+\frac{2}{9\nu_2}\left(\frac{\nu_1 F'}{\nu_1+\lambda}\right)^{\frac{2}{3}}\right]^{\frac{1}{2}}}. \tag{17}$$

Equation (16) was used to compute the 'new approximation' values given in Table 5.

Note that when $\lambda = 0$, the above approximation reduces to the approximation for the (central) F distribution given by Paulson (1942). We also point out that approximation (17) produced much more accurate results than either of Johnson's approximations (cf. Johnson, 1959). The latter are based on normal approximations to the difference between a negative binomial variable and an independent Poisson variable; this difference Johnson shows is related to the non-central F variable.

We would like to thank Miss Shirely Young, D. Liepmann, and Mrs Marion T. Carson for performing the extensive calculations.

REFERENCES

ABDEL-ATY, S. H. (1954). Approximate formulae for the percentage points and the probability integral of the non-central χ^2 distribution. *Biometrika*, **41**, 538–42.

CAMPBELL, G. A. (1923). Probability curves showing Poisson's exponential summation. *Bell Syst. Tech. J.* **2**, 95–113.

JOHNSON, N. L. (1959). On an extension of the connection between Poisson and χ^2 distributions. *Biometrika*, **46**, 352–63.

MERRINGTON, M. (1941). Numerical approximations to the percentage points of the χ^2 distribution. *Biometrika*, **32**, 200–2.

PATNAIK, P. B. (1949). The non-central χ^2- and F-distributions and their applications. *Biometrika*, **36**, 203–32.

PAULSON, E. (1942). An approximate normalization of the analysis of variance distribution. *Ann. Math. Statist.* **13**, 233–5.

PEARSON, E. S. & HARTLEY, H. O. (1954). *Biometrika Tables for Statisticians*, **1**. Cambridge University Press.

[*Editorial Note.* An alternative method of approximation to the distribution of non-central F is given by Barton, David & O'Neill in the paper which follows. On pp. 430–1 below some comparisons of the two approximations have been made. E.S.P.]

Biometrika (1960), **47**, 3 *and* 4, *p.* 417
Printed in Great Britain

Some properties of the distribution of the logarithm of non-central *F*

BY D. E. BARTON, F. N. DAVID† AND ANNE F. O'NEILL
University College, London and Wheaton College, Norton, U.S.A.

1. INTRODUCTION

Given two independent variables, χ'^2 distributed as a non-central chi-square with degrees of freedom f_1 and non-central parameter λ, and χ^2 distributed as a central chi-square with degrees of freedom f_2, then we define non-central F as

$$F^* = \left(\frac{\chi'^2/f_1}{\chi^2/f_2}\right).$$

The distribution function of F^* is

$$p(F^*) = \sum_{j=0}^{\infty}\left[\frac{e^{-\frac{1}{2}\lambda}(\frac{1}{2}\lambda)^j}{j!}\frac{(\frac{1}{2}(f_1+f_2)+j-1)!}{(\frac{1}{2}f_1+j-1)!\,(\frac{1}{2}f_2-1)!}\left(\frac{f_1}{f_2}\right)^{\frac{1}{2}f_1+j}F^{*\frac{1}{2}f_1-1+j}\left(1+\frac{f_1}{f_2}F^*\right)^{-\frac{1}{2}(f_1+f_2)-j}\right].$$

This distribution was first given by Fisher (1928). In a different functional form, i.e.

$$E^2 = \frac{f_1F^*}{f_2+f_1F^*}$$

it was studied by Wishart (1932) and by Tang (1938) who gave a few values of the probability integral. Patnaik (1949) approximated to the distribution of F^* by a central F distribution. It is clear that any tabulation of the distribution of F^* will be a formidable proceeding. There are three parameters f_1, f_2 and λ which intervene explicitly. If we are concerned with the power function of the F-test there must also be the values of the central F at which $p(F^*)$ will need to be evaluated. We discuss here the log-transformation of F^* and the possibility of a normal distribution as an approximating function to the transformed variable.§

2. THE CUMULANTS OF $\log F^*$

The distributions of F and F^* are extremely skew and it does not appear profitable to consider them directly. Fisher & Cornish (1937) showed that expansions using

$$z = \tfrac{1}{2}\log F$$

could be obtained which gave accurate arithmetical results for the probability integral of z (and therefore of F). Accordingly we look at

$$z^* = \tfrac{1}{2}\log F^*,$$

and consider the case first when λ, f_1 and f_2 are small. Since the distribution of z is not very different from normal and can be expressed as a series of Hermite polynomials weighted by the standardized cumulants of z, it is reasonable to suppose that for λ small the same will be true of the distribution of z^*. Wishart (1947) gave expansions for the moments of z but these forms of expansion are not convenient for our purpose. We have that

$$2z^* = \log(f_2/f_1)+\log(\chi'^2)-\log(\chi^2).$$

† With the partial support of the office of Naval Research while at the University of California, Berkeley.

§ Some comparisons of this approximation with that suggested by Severo & Zelen in the preceding paper are made in an Editorial Note on pp. 430–1 below.

For ease of writing let $a_1 = \frac{1}{2}f_1, \quad \theta = \frac{1}{2}it.$

The characteristic function of $\log(\frac{1}{2}\chi'^2)$ may be written as

$$\phi_\lambda(t) = \sum_{r=0}^{\infty} \frac{(\frac{1}{2}\lambda)^r}{r!} \Delta^r \left(\frac{\Gamma(a_1+\theta)}{\Gamma(a_1)} \right), \tag{1}$$

where

$$\Delta^r \left(\frac{\Gamma(a_1+\theta)}{\Gamma(a_1)} \right) = \sum_{j=0}^{r} (-1)^{r-j} {}^rC_j \frac{\Gamma(a_1+j+\theta)}{\Gamma(a_1+j)}. \tag{2}$$

By manipulation of the factorial expressions under the right-hand sum it may be shown that this equation reduces to

$$\Delta^r \left(\frac{\Gamma(a_1+\theta)}{\Gamma(a_1)} \right) = \frac{\Gamma(a_1+\theta)}{\Gamma(a_1)} \frac{\theta^{(r)}}{a_1^{[r]}}, \tag{3}$$

where $\theta^{(r)} = \theta(\theta-1)\dots(\theta-r+1)$

and $a_1^{[r]} = a_1(a_1+1)\dots(a_1+r-1).$

Consequently the characteristic function of $\log(\frac{1}{2}\chi'^2)$ may be written as

$$\phi_\lambda(t) = \frac{\Gamma(a_1+\theta)}{\Gamma(a_1)} \sum_{r=0}^{\infty} \left(\frac{\lambda}{2} \right)^r \frac{1}{r!} \frac{\theta^{(r)}}{a_1^{[r]}}, \tag{4}$$

while the value for the central $\log(\frac{1}{2}\chi^2)$ is

$$\phi_0(t) = \frac{\Gamma(a_1+\theta)}{\Gamma(a_1)}. \tag{5}$$

We follow the notation of Legendre and Gauss (cf. H. T. Davis's (1933–5) Tables) and write

$$\psi(x) = \frac{d}{dx}\log\Gamma(x), \quad \psi'(x) = \frac{d}{dx}\psi(x)$$

and so on. Taking the logarithm of the characteristic function, differentiating with respect to θ, and then equating θ to zero as many cumulants can be obtained as desired. Let $\kappa_{r1}(\lambda)$ denote the rth cumulant of $\log(\frac{1}{2}\chi'^2)$ and κ_{r2} the rth cumulant of $\log(\frac{1}{2}\chi^2)$. Then

$$\left.\begin{aligned}
\kappa_{11}(\lambda) &= \psi(a_1) + M_1, \\
\kappa_{21}(\lambda) &= \psi'(a_1) + M_2 - M_1^2, \\
\kappa_{31}(\lambda) &= \psi''(a_1) + M_3 - 3M_2M_1^2 + 2M_1^3, \\
\kappa_{41}(\lambda) &= \psi'''(a_1) + (M_4 - 4M_1M_3 + 6M_1^2M_2 - 3M_1^4) - 3\{M_2 - M_1^2\}^2,
\end{aligned}\right\} \tag{6}$$

where

$$M_v = v! \sum_{r=v}^{\infty} \frac{(\frac{1}{2}\lambda)^r (-1)^{r-v}}{r!\, a_1^{[r]}} |S_r^v|. \tag{7}$$

$|S_r^v|$ is Stirling's number of the first kind in modulus† and, conventionally, $|S_r^1|$ is $(r-1)!$ It will be noticed that we may write

$$\phi_\lambda(t) = \phi_0(t) \sum_j \frac{\theta^j M_j}{j!} \tag{8}$$

so that formally $\log(\frac{1}{2}\chi'^2)$ is the sum of two independent random variables, $\log(\frac{1}{2}\chi^2)$ plus a 'random variable' with crude moments M_j. The quantities containing the M's in the

† These will be given in Barton, David & Kendall's *Symmetric Function and Allied Tables* at present in preparation for publication.

expression for $\kappa_{r1}(\lambda)$ are the cumulant functions of the 'moments' M_j. If we denote these by $\mathfrak{K}_r$ we have

$$\left.\begin{aligned}\kappa_1(2z^*) &= \psi(a_1)-\psi(a_2)+\log(a_2/a_1)+\mathfrak{K}_1,\\ \kappa_2(2z^*) &= \psi'(a_1)+\psi'(a_2)+\mathfrak{K}_2,\\ \kappa_3(2z^*) &= \psi''(a_1)-\psi''(a_2)+\mathfrak{K}_3,\\ \kappa_4(2z^*) &= \psi'''(a_1)+\psi'''(a_2)+\mathfrak{K}_4,\end{aligned}\right\} \tag{9}$$

where $a_1 = \frac{1}{2}f_1$, $a_2 = \frac{1}{2}f_2$. As many more cumulants may be written down as desired. $\psi(a)$, $\psi'(a)$, $\psi''(a)$, $\psi'''(a)$, and $\mathfrak{K}_1$, $\mathfrak{K}_2$, $\mathfrak{K}_3$, $\mathfrak{K}_4$ are given in Tables A and B, respectively, printed on pp. 425–29.

A simple recurrence relation may be found for these M's. Let

$$G_{a_1} = \sum_{r=0}^{\infty}\left(\frac{\lambda}{2}\right)^r \frac{1}{r!}\frac{\theta^{(r)}}{a_1^{[r]}} = \sum_j \frac{\theta^j M_j}{j!},$$

$$\begin{aligned}(\theta+1)G_{a_1} &= \sum_{r=0}\left(\frac{\lambda}{2}\right)^r \frac{1}{r!\,a_1^{[r]}}[\theta^{(r+1)}+(r+1)\theta^{(r)}]\\ &= G_{a_1}+\frac{2(a_1-1)(a_1-2)}{\lambda}[G_{a_1-2}-G_{a_1-1}]+(a_1-1)[G_{a_1-1}-G_{a_1}],\end{aligned}$$

so that

$$\theta G_{a_1} = \frac{2(a_1-1)(a_1-2)}{\lambda}[G_{a_1-2}-G_{a_1-1}]+(a_1-1)[G_{a_1-1}-G_{a_1}]. \tag{10}$$

Thus

$$rM_{r-1}(a_1+2) = \frac{2a_1(a_1+1)}{\lambda}(M_r(a_1)-M_r(a_1+1))+(a_1+1)(M_r(a_1+1)-M_r(a_1+2)).$$

This relation may be used to build up tables of M_r for higher values of a_1, in terms of M_{r-1} and initial values of M_r. The series for M_r are rapidly convergent even for $\lambda = 10$ and $f_1 = 20$, while for $\lambda = 1$ and $f_1 = 2$, ($a_1 = 1$), only three terms are required for seven decimal place accuracy.

It is clear that the series (7) given above will be useful for most practical purposes and it is this which we have used for tabulation. For very large λ the calculations may be lengthy and for this case we would propose an alternative expansion. We have

$$\begin{aligned}2z^* &= \log\frac{f_2}{f_1}-\log\chi^2+\log(\chi'^2-(f_1+\lambda)+f_1+\lambda)\\ &= \log\left(\frac{f_2(f_1+\lambda)}{f_1}\right)-\log(\chi^2)+\log\left(1+\frac{\chi'^2-(f_1+\lambda)}{f_1+\lambda}\right).\end{aligned}$$

The moments of $\log(\frac{1}{2}\chi^2)$ are known exactly and are as given above (i.e. they are the leading terms in (6)). For brevity write

$$H = \log\left(1+\frac{\chi'^2-(f_1+\lambda)}{f_1+\lambda}\right). \tag{11}$$

We expand the right-hand side and using the fact that the rth cumulant of χ'^2 is

$$2^{r-1}(r-1)!\,(f_1+r\lambda)$$

obtain

$$\left.\begin{aligned}
\kappa_1(H) &= \left[-\frac{1}{f_1+\lambda}-\frac{\lambda}{(f_1+\lambda)^2}\right]+\left[-\frac{1}{3(f_1+\lambda)^2}-\frac{2\lambda}{(f_1+\lambda)^3}-\frac{3\lambda^2}{(f_1+\lambda)^4}\right]+\left[\frac{4\lambda^2}{(f_1+\lambda)^5}-\frac{20\lambda^3}{(f_1+\lambda)^6}\right]+\ldots,\\
\kappa_2(H) &= \left[\frac{2}{f_1+\lambda}+\frac{2\lambda}{(f_1+\lambda)^2}\right]+\left[\frac{2}{(f_1+\lambda)^2}+\frac{4\lambda}{(f_1+\lambda)^3}+\frac{10\lambda^2}{(f_1+\lambda)^4}\right]+\left[\frac{4}{3(f_1+\lambda)^3}+\frac{4\lambda}{(f_1+\lambda)^4}+\frac{256\lambda^3}{3(f_1+\lambda)^6}\right]+\ldots,\\
\kappa_3(H) &= \left[-\frac{4}{(f_1+\lambda)^2}-\frac{8\lambda}{(f_1+\lambda)^3}-\frac{12\lambda^2}{(f_1+\lambda)^4}\right]+\left[-\frac{8}{(f_1+\lambda)^3}-\frac{24\lambda}{(f_1+\lambda)^4}-\frac{48\lambda^2}{(f_1+\lambda)^5}-\frac{176\lambda^3}{(f_1+\lambda)^6}\right]+\ldots,\\
\kappa_4(H) &= \left[\frac{16}{(f_1+\lambda)^3}+\frac{48\lambda}{(f_1+\lambda)^4}+\frac{96\lambda^2}{(f_1+\lambda)^5}+\frac{160\lambda^3}{(f_1+\lambda)^6}\right]+\ldots.
\end{aligned}\right\} \tag{12}$$

Because of the independence of χ^2 and χ'^2 we have

$$\kappa_1(2z^*) = \log\left(f_2(f_1+\lambda)/f_1\right)-\kappa_{12}+\kappa_1(H),$$
$$\kappa_r(2z^*) = \kappa_r(H)+(-1)^r\kappa_{r2} \quad (r \geqslant 2).$$

3. Evaluation of the power function of the F-test

Fisher & Cornish (1937) showed that an Edgeworth expansion was suitable for calculating the probability integral of z provided the degrees of freedom of numerator and denominator are large. Following a procedure indicated by David (1959), we shall use an Edgeworth expansion and the moments of $2z^*$ to evaluate the power function. Our procedure will depend not directly on the degrees of freedom of numerator and denominator but on the relative values of the moments of central $2z$ and on the corrective term to be added in order to obtain the moments of $2z^*$. Write

$$\left.\begin{aligned}
\kappa_1(2z)-m &= \sigma l_1, & \kappa_1(2z^*)-m &= \sigma L_1,\\
\kappa_2(2z)-\sigma^2 &= \sigma^2 l_2, & \kappa_2(2z^*)-\sigma^2 &= \sigma^2 L_2,\\
\kappa_r(2z) &= \sigma^r l_r, & \kappa_r(2z^*) &= \sigma^r L_r.
\end{aligned}\right\} \tag{13}$$

m and σ^2 are at choice and we will choose them so that l_1 and l_2 are zero. If

$$\frac{2z-\kappa_1(2z)}{\sqrt{\{\kappa_2(2z)\}}} = \frac{2z-m}{\sigma} = x$$

and

$$g(x) = \frac{1}{\sqrt{(2\pi)}}e^{-\frac{1}{2}x^2},$$

then

$$P\left\{\frac{2z-m}{\sigma} < x_\alpha\right\} = \int_{-\infty}^{x_\alpha} g(x)\,dx + g(x_\alpha)\left[-\tfrac{1}{6}l_3H_2-\tfrac{1}{24}l_4H_3-\tfrac{1}{72}l_3^2H_5\right]+\ldots$$
$$= 1-\alpha, \tag{14}$$

where H_r stands for the rth Hermite polynomial evaluated at the point x_α. Again

$$P\{2z^* < \sigma x_\alpha + m\} = \int_{-\infty}^{x_\alpha} g(x)\,dx$$
$$+ g(x_\alpha)\left[(-L_1-\tfrac{1}{6}L_3H_2)-(\tfrac{1}{2}L_1^2H_1+\tfrac{1}{2}L_2H_1+\tfrac{1}{6}L_1L_3H_3+\tfrac{1}{24}L_4H_3+\tfrac{1}{72}L_3^2H_5)-\ldots\right] \tag{15}$$

so that the amount by which $P\{z < z_\alpha\}$ is exceeded in the z^* distribution is approximately

$$g(x_\alpha)\Big[-\frac{(\kappa_1(2z^*)-\kappa_1(2z))}{\sigma_{2z}}-\tfrac{1}{6}H_2\frac{(\kappa_3(2z^*)-\kappa_3(2z))}{\sigma_{2z}^3}-\tfrac{1}{2}H_1\Big(\frac{\kappa_1(2z^*)-\kappa_1(2z)}{\sigma_{2z}}\Big)^2$$
$$-\tfrac{1}{2}H_1\Big(\frac{\kappa_2(2z^*)-\kappa_2(2z)}{\sigma_{2z}^2}\Big)-\tfrac{1}{6}H_3\frac{\kappa_3(2z^*)}{\sigma_{2z}^3}\Big(\frac{\kappa_1(2z^*)-\kappa_1(2z)}{\sigma_{2z}}\Big)$$
$$-\tfrac{1}{24}H_3\Big(\frac{\kappa_4(2z^*)-\kappa_4(2z)}{\sigma_{2z}^4}\Big)-\tfrac{1}{72}H_5\Big(\frac{\kappa_3^2(2z^*)-\kappa_3^2(2z)}{\sigma_{2z}^6}\Big)\Big]$$
$$=g(x_\alpha)\Big[-\frac{\mathfrak{K}_1}{\sigma_{2z}}-\tfrac{1}{6}H_2\frac{\mathfrak{K}_3}{\sigma_{2z}^3}-\tfrac{1}{2}H_1\Big(\frac{\mathfrak{K}_1}{\sigma_{2z}}\Big)^2-\tfrac{1}{2}H_1\frac{\mathfrak{K}_2}{\sigma_{2z}^2}-\tfrac{1}{6}H_3\frac{\kappa_3(2z^*)}{\sigma_{2z}^3}\frac{\mathfrak{K}_1}{\sigma_{2z}}$$
$$-\tfrac{1}{24}H_3\frac{\mathfrak{K}_4}{\sigma_{2z}^4}-\tfrac{1}{72}H_5\frac{(\kappa_3(2z^*)+\kappa_3(2z))}{\sigma_{2z}^3}\frac{\mathfrak{K}_3}{\sigma_{2z}^3}\Big]. \qquad (16)$$

Further terms depending on the higher cumulants can be added if required. This expansion clearly depends on known values of z_α. These have been given directly by Fisher & Yates (1953, Table V) for four values of α. Since $z_\alpha = \frac{1}{2}\log F_\alpha$, values for further levels can be found from Merrington & Thompson's tables (1943). If the power of the F-test is required using a significance level of z or F which has not been tabulated, it will be necessary either to interpolate between known percentage points or to use the inverse Edgeworth expansion to determine such points. Thus if ξ is the solution of the equation

$$\int_{-\infty}^{\xi} g(x)\,dx = 1-\alpha,$$

then the central z_α can be determined from,

$$z_\alpha = \kappa_1(z)+\sigma_z\{\xi+\tfrac{1}{6}l_3H_2(\xi)+\tfrac{1}{24}l_4H_3(z)-\tfrac{1}{36}l_3^2(2H_3(\xi)+\xi)+\ldots\},$$

where the l's retain their previous sense and $H_r(\xi)$ is the rth Hermite polynomial evaluated at ξ. Fisher & Cornish gave a table of these Hermites for some values of α. Once z_α is found the previous expansion may be used.

Under certain conditions, for example, when the distribution of z^* is far from that of z, the series given by (16) will not converge fast enough to be useful. It is also possible that it may be desired to tabulate the distribution of z^* without reference to the percentage points of z. In such cases the Edgeworth expansion may be used directly with the standardized cumulants of z^* or $2z^*$. If

$$\kappa_1(2z^*)-m^* = \mathscr{L}_1\sigma^*, \quad \kappa_2(2z^*)-\sigma^{*2} = \mathscr{L}_2\sigma^{*2}, \quad \kappa_r(2z^*) = \sigma^{*r}\mathscr{L}_r \qquad (17)$$

and we choose m^* and σ^{*2} equal to $\kappa_1(2z^*)$ and $\kappa_2(2z^*)$, respectively, so that $\mathscr{L}_1$ and $\mathscr{L}_2$ are zero, then

$$P\{z^* < Z\} = \int_{-\infty}^{X} g(x)\,dx + g(X)\{-\tfrac{1}{6}H_2(X)\mathscr{L}_3-\tfrac{1}{24}H_3(X)\mathscr{L}_4-\tfrac{1}{72}H_5(X)\mathscr{L}_3^2+\ldots\},$$

where

$$X = \frac{2Z-\kappa_1(2z^*)}{(\kappa_2(2z^*))^{\frac{1}{2}}}. \qquad (18)$$

The inverse of the expansions follow directly. Let ξ be the unit normal deviate defined by

$$\int_{-\infty}^{\xi} g(x)\,dx = 1-\alpha.$$

Then $$Y^*-\xi = \tfrac{1}{6}\mathscr{L}_3(\xi^2-1)+\tfrac{1}{24}\mathscr{L}_4(\xi^3-3\xi)-\tfrac{1}{36}\mathscr{L}_3^2(2\xi^2-5\xi)+\ldots,$$

where $\mathscr{L}_r$ is as defined in (17) and m^* and σ^{*2} are equal to $\kappa_1(2z^*)$ and $\kappa_2(2z^*)$ as before. The value of $2z^*$ which cuts the distribution into the proportions $1-\alpha:\alpha$ will be

$$Y^*\sqrt{\{\kappa_2(2z^*)\}}+\kappa_1(2z^*).$$

An alternative expansion in terms of the central distribution of $2z$ is also possible, using the standardized l_r and L_r. If Y be the value of $2z$ which divides the distribution of central $2z$ in the ratio $1-\alpha:\alpha$ then the value of Y^* which divides the non-central distribution in this ratio is

$$Y^*-Y = +\frac{\mathfrak{K}_1}{\sigma}+\tfrac{1}{6}(\xi^2-1)\frac{\mathfrak{K}_3}{\sigma^3}+\tfrac{1}{24}\xi\frac{\mathfrak{K}_2}{\sigma^2}+\tfrac{1}{24}(\xi^3-3\xi)\frac{\mathfrak{K}_4}{\sigma^4}-\tfrac{1}{36}(2\xi^3-5\xi)\left(\frac{\mathfrak{K}_3}{\sigma^3}\right)\left(2+\frac{\mathfrak{K}_3}{\sigma^3}\right)+\ldots,$$

where $$\sigma = \sqrt{\{\kappa_2(2z)\}}.$$

This gives a standardized deviation which must be multiplied by σ. This second procedure will be useful for small values of the non-central parameter when the differences between the central and the non-central distributions are not large.

4. Momental ratios for the distribution of z^*

The distribution of z^* is not too far from the normal, the largest discrepancies occurring when λ is large and $f_1 = 2a_1$, the degrees of freedom of the numerator, and $f_2 = 2a_2$, the degrees of freedom of the denominator, are very different. For interest we give in Table 1 below a few selected values.

Table 1. *Momental ratios, $\sqrt{\beta_1}$ and β_2-3 of $z^* = \frac{1}{2}\log F^*$ for selected values of f_1, f_2 and λ*

		$\sqrt{\beta_1}$, f_2				β_2-3, f_2			
λ	f_1	2	4	10	20	2	4	10	20
0	2	0	−0·5772	−0·9238	−1·0336	1·2	1·3327	1·8707	2·1210
½		0·0014	− ·5850	− ·9406	−1·0540	1·2192	1·3634	1·9239	2·1855
1		·0088	− ·5977	− ·9741	−1·0958	1·2647	1·4357	2·0526	2·3430
3		·1165	− ·5835	−1·0787	−1·2533	1·4884	1·7686	2·7550	3·2616
0	4	0·5772	0	−0·4407	−0·6050	1·3327	0·5938	0·6868	0·8820
½		·5790	0·0004	− ·4436	− ·6098	1·3380	·5980	·6940	·8928
1		·5845	·0025	− ·4490	− ·6198	1·3517	·6084	·7118	·9195
5		·6901	·0985	− ·4223	− ·6453	1·5420	·7025	·8343	1·1501
0	10	0·9238	0·4407	0	−0·2023	1·8707	0·6868	0·2187	0·2228
½		·9241	·4410	0·0000	− ·2026	1·8714	·6874	·2191	·2233
1		·9249	·4419	·0003	− ·2033	1·8736	·6891	·2202	·2248
0	20	1·0336	0·6050	0·2023	0	2·1210	0·8820	0·2228	0·1049
½		1·0337	·6051	·2024	0·0000	2·1211	·8821	·2229	·1049
1		1·0338	·6053	·2025	·0000	2·1215	·8825	·2230	·1050

5. Numerical examination of the approximations

Tang (1938) computed some exact values of the probability distribution function of F^*, or rather of the related E^*. These did not prove altogether suitable for our present purposes which were to test our approximate series for small values of λ. Exact values of the probability integral of F^* over the range which we required were therefore computed *de novo*

Table 2. *Comparison of approximations to* $P\{z^* < z_\alpha\}$

	$\lambda = \frac{1}{2}, f_1 = f_2 = 10, \beta_1 = 0{\cdot}0000, \beta_2 = 3{\cdot}2191$				$\lambda = \frac{1}{2}, f_1 = 10, f_2 = 20, \beta_1 = 0{\cdot}0410, \beta_2 = 3{\cdot}2233$			
		Error in approximations				Error in approximations		
α	Exact value	Normal	Series I	Series II	Exact value	Normal	Series I	Series II
0·001	0·9988	0·0005	0·0000	0·0000	0·9987	−0·0009	−0·0002	−0·0002
·005	·9940	11	1	−1	·9935	−25	− 5	2
·01	·9881	12	0	−2	·9873	−38	− 5	4
·025	·9708	5	−2	−2	·9692	−57	− 8	5
·05	·9423	− 8	−3	1	·9398	−65	− 1	6
·10	·8868	−31	−5	2	·8830	−53	33	3
·25	·7258	−45	1	2	·7205	+35	12	−2
·50	·4699	7	7	−1	·4652	133	6	−5
·75	·2268	47	1	−3	·2243	81	−10	−6
·90	·0878	22	−5	−2	·0869	−24	− 6	1
·95	·0431	1	−4	0	·0426	−51	1	6
·975	·0212	− 9	−1	2	·0210	−51	4	6
·99	·0084	−12	0	1	·0083	−35	4	3
·995	·0041	− 9	1	1	·0041	−23	4	1
·999	·0008	− 4	1	0	·0008	− 6	1	−1

	$\lambda = 1, f_1 = f_2 = 10, \beta_1 = 0{\cdot}0000, \beta_2 = 3{\cdot}2202$				$\lambda = 1, f_1 = 10, f_2 = 20, \beta_1 = 0{\cdot}0413, \beta_2 = 3{\cdot}2248$			
		Error in approximations (approx. − true)				Error in approximations (approx. − true)		
α	Exact value	Normal	Series I	Series II	Exact value	Normal	Series I	Series II
0·001	0·9986	0·0005	0·0001	−0·0001	0·9983	−0·0011	0·0004	−0·0001
·005	·9929	11	1	−2	·9918	−29	−10	1
·01	·9861	11	−1	−2	·9843	−43	− 9	4
·025	·9662	3	−4	−1	·9627	−61	−14	6
·05	·9342	−12	−7	1	·9288	−65	0	5
·10	·8730	−35	−8	3	·8658	−52	7	−4
·25	·7015	−43	3	3	·6910	51	24	−3
·50	·4411	14	13	−1	·4321	136	9	−8
·75	·2056	46	0	−4	·2009	67	−19	−5
·90	·0770	18	−8	−1	·0754	−32	−10	2
·95	·0371	− 2	−6	1	·0364	−54	2	5
·975	·0180	−11	−2	2	·0177	−49	8	5
·99	·0070	−12	0	1	·0069	−32	8	4
·995	·0034	− 9	2	1	·0034	−20	7	0
·999	·0007	− 4	1	−1	·0007	− 6	2	−2

			Error in approximations (approx. − true)					Error in approximations (approx. − true)				
		Exact value	Normal	Series I	Series II	Series III	S_u	Exact value	Normal	Series I	Series II	Series III	S_u
		$\lambda = 3, \alpha = 0{\cdot}01$						$\lambda = 3, \alpha = 0{\cdot}05$					
$f_1 = 2$	f_2												
	2	0·975	0·007	−0·001	−0·004	−0·003	0·000	0·881	−0·015	0·016	0·010	0·006	0·000
	4	·957	−16	−19	+11	+1	1	·824	−19	−3	4	11	−1
	10	·920	−37	−28	10	16	1	·752	− 1	34	− 6	8	−3
	20	·895	−38	−22	3	16	0	·718	12	55	−10	5	−6
		$\lambda = 5, \alpha = 0{\cdot}01$						$\lambda = 5, \alpha = 0{\cdot}05$					
$f_1 = 4$	f_2												
	2	0·978	0·012	0·005	−0·006	0·001	0·000	0·892	−0·004	−0·013	0·004	0·002	−0·001
	4	·960	3	0	0	0	0	·833	−12	−7	4	2	0
	10	·916	−14	−6	4	2	1	·738	0	23	−1	3	0
	20	·878	−16	−1	5	4	1	·681	15	48	−2	3	0

from Tang's formulae.† For the smaller values of λ, three approximations for $P\{z^* < z_\alpha\}$ are compared in the upper half of Table 2, using values of 'central' z_α taken from the two tables referred to above:

(i) the crude normal approximation, obtained by assuming that $\{z^* - \kappa_1(z^*)\}/\sqrt{\kappa_2(z^*)}$ is a unit normal deviate;

(ii) Series I approximation from equation (16);

(iii) Series II approximation from equation (18).

The upper half of the table gives the differences, in units of the fourth decimal place, between the approximate and true value. It will be seen that the crude normal approximation is in some cases surprisingly reasonable.

Some calculations were carried out to see what happened for larger λ and for f_1 and f_2 either small or widely different. The results are shown in the lower half of the table. In these cases Series I is improved by adding further terms of the Hermite expansion to equation (16); we have termed this Series III. It was also found that using the correct first four moments of z^* and the functional form called S_u by Johnson (1949), the probability integral of z^* could be very closely reproduced, at any rate for the points with $\alpha = 0{\cdot}05$ and $0{\cdot}01$. The labour of fitting an S_u curve lies in finding its parameters, given the moments; this labour is compensated to some extent by the fact that once fitted, an approximation can be found quickly for the whole of the probability integral of z^*. But the procedure was thought to be too lengthy to warrant its being put to general use and we therefore abandoned it.

The Table B for calculating the moments of z^* has been computed for $\lambda = 0{\cdot}0\,(0{\cdot}5)\,4, 5$. It is assumed that for intermediate values of λ either the series expansion will be used to calculate the M's, and therefore the $\mathfrak{K}$'s, or that there will be interpolation into the existing tables. In order to examine this latter point, the values of $\mathfrak{K}_1$, $\mathfrak{K}_2$, $\mathfrak{K}_3$ and $\mathfrak{K}_4$ for $\lambda = 1{\cdot}625$, $f_1 = 4$ were obtained from the series for M and also by 4-point Lagrangian interpolation, with the following results:

	$\mathfrak{K}_1$	$\mathfrak{K}_2$	$\mathfrak{K}_3$	$\mathfrak{K}_4$
From series	0·35788	−0·03749	0·01211	−0·00635
By interpolation	·35787	− ·03743	·01203	− ·00627

The accuracy obtained by interpolation is here clearly adequate for the application of equations (16) or (17).

REFERENCES

DAVID, F. N. (1959). *Biometrika*, **46**, 123–9.

DAVIS, H. T. (1933–5). *Tables of Higher Mathematical Functions*, **1** and **2**. Indiana: Principia Press.

FISHER, R. A. (1928). *Proc. Roy. Soc.* A, **121**, 654–73.

FISHER, R. A. & CORNISH, E. A. (1937). *Rev. Int. Statist. Inst.* **5**, 307–21.

FISHER, R. A. & YATES, F. (1953). *Statistical Tables for Biological, Agriculture and Medical Research*. 4th Edition. Edinburgh: Oliver and Boyd.

JOHNSON, N. L. (1949). *Biometrika*, **36**, 149–76.

MERRINGTON, M. & THOMPSON, C. M. (1943). *Biometrika*, **33**, 73–88.

PATNAIK, P. B. (1949). *Biometrika*, **36**, 202–32.

TANG, P. C. (1938). *Stat. Res. Mem.* **2**, 126–49.

WISHART, J. (1932). *Biometrika*, **24**, 441–56.

WISHART, J. (1947). *Biometrika*, **34**, 170–8.

† λ as used by us is one half of the λ used in Tang.

Table A. *Values of the logarithmic derivatives of the gamma function*

$a = \frac{1}{2}f$	$\psi(a)$	$\psi'(a)$	$\psi''(a)$	$\psi'''(a)$
1	−0·5772 1566	1·6449 3407	−2·4041 1381	6·4939 3940
1·5	+0·0364 8997	0·9348 0220	−0·8287 9664	1·4090 9103
2	0·4227 8434	·6449 3407	− ·4041 1381	·4939 3940
2·5	0·7031 5664	0·4903 5776	−0·2362 0405	0·2239 0585
3	0·9227 8434	·3949 3407	− ·1541 1381	·1189 3940
3·5	1·1031 5664	·3303 5776	− ·1082 0405	·0703 0585
4	1·2561 1767	·2838 2296	− ·0800 3973	·0448 6533
4·5	1·3888 7093	·2487 2510	− ·0615 5682	·0303 2251
5	1·5061 1767	0·2213 2296	−0·0487 8973	0·0214 2783
5·5	1·6110 9315	·1993 4239	− ·0396 0895	·0156 9059
6	1·7061 1767	·1813 2296	− ·0327 8973	·0111 8278
6·5	1·7929 1133	·1662 8454	− ·0275 8791	·0091 3366
7	1·8727 8434	·1535 4518	− ·0235 3047	·0071 9820
7·5	1·9467 5748	0·1426 1590	−0·0203 0525	0·0057 7244
8	2·0156 4148	·1331 3701	− ·0176 9957	·0046 9924
8·5	2·0800 9082	·1248 3812	− ·0155 6451	·0038 7614
9	2·1406 4148	·1175 1201	− ·0137 9332	·0032 3440
9·5	2·1977 3788	·1109 9729	− ·0123 0785	·0027 2673
10	2·2517 5259	0·1051 6634	−0·0110 4984	0·0023 1990
10·5	2·3030 0103	·0999 1696	− ·0099 7514	·0019 9001
11	2·3517 5259	·0951 6634	− ·0090 4984	·0017 1990
11·5	2·3982 3913	·0908 4666	− ·0082 4747	·0014 9646
12	2·4426 6168	·0869 0187	− ·0075 4721	·0013 1009
12·5	2·4851 9565	0·0832 8522	−0·0069 3244	0·0011 5341
13	2·5259 9501	·0799 5743	− ·0063 8980	·0010 2074
13·5	2·5651 9565	·0768 8522	− ·0059 0844	·0009 0765
14	2·6029 1809	·0740 4027	− ·0054 7947	·0008 1066
14·5	2·6392 6973	·0713 9826	− ·0050 9555	·0007 2701
15	2·6743 4666	0·0689 3823	−0·0047 5060	0·0006 5448
15·5	2·7082 3524	·0666 4201	− ·0044 3952	·0005 9128

Note. $\psi(a) = \frac{d}{da}\log\Gamma(a)$, $\psi'(a) = \frac{d}{da}\psi(a)$, etc.

Table B. *The $\mathfrak{R}$-functions of equation* (9)

(1) Values of $\mathfrak{R}_1 = M_1$

f_1 \ λ	0·5	1·0	1·5	2·0	2·5	3·0	3·5	4·0	5·0
2	0·235 204	0·443 842	0·629 874	0·796 600	0·946 773	1·082 700	1·206 320	1·319 263	1·518 421
3	·158 714	·302 928	·434 508	·555 042	·665 885$^+$	0·768 195$^+$	0·862 967	0·951 055$^-$	1·110 030
4	·120 001	·230 781	·333 386	·428 720	·517 569	·600 614	·678 449	·751 596	0·885 587
5	0·096 556	0·186 705$^-$	0·271 094	0·350 296	0·424 810	0·495 079	0·561 491	0·624 391	0·740 847
6	·080 813	·156 903	·228 704	·296 600	·360 932	·422 005$^-$	·480 090	·535 430	·638 721
7	·069 503	·135 375$^+$	·197 922	·257 415$^+$	·314 100	·368 195$^+$	·419 899	·469 389	·562 357
8	·060 980	·119 080	·174 522	·227 507	·278 218	·326 817	·373 453	·418 262	·502 881
9	·054 325$^-$	·106 307	·156 116	·203 903	·249 807	·293 954	·336 459	·377 427	·455 132
10	0·048 983	0·096 023	0·141 251	0·184 784	0·226 732	0·267 193	0·306 259	0·344 013	0·415 889
11	·044 599	·087 561	·128 988	·168 974	·207 606	·244 962	·281 117	·316 138	·383 024
12	·040 938	·080 476	·118 697	·155 677	·191 486	·226 189	·259 846	·292 512	·355 076
13	·037 833	·074 456	·109 936	·144 335$^+$	·177 712	·210 119	·241 607	·272 221	·331 001
14	·035 166	·069 276	·102 385$^-$	·134 544	·165 801	·196 202	·225 788	·254 599	·310 036
15	0·032 852	0·064 772	0·095 809	0·126 004	0·155 398	0·184 030	0·211 935$^-$	0·239 145$^+$	0·291 607
16	·030 823	·060 820	·090 030	·118 489	·146 232	·173 292	·199 699	·225 481	·275 275$^-$
17	·029 031	·057 323	·084 910	·111 824	·138 094	·163 747	·188 810	·213 308	·260 698
18	·027 436	·054 207	·080 344	·105 872	·130 818	·155 206	·179 057	·202 394	·247 604
19	·026 007	·051 414	·076 245$^+$	·100 525$^-$	·124 275$^-$	·147 516	·170 269	·192 552	·235 777
20	0·024 720	0·048 894	0·072 545$^+$	0·095 693	0·118 358	0·140 557	0·162 309	0·183 629	0·225 039
21	·023 554	·046 611	·069 189	·091 307	·112 981	·134 228	·155 064	·175 503	·215 245$^-$
22	·022 494	·044 531	·066 130	·087 306	·108 073	·128 448	·148 442	·168 070	·206 275$^+$
23	·021 524	·042 630	·063 331	·083 642	·103 576	·123 147	·142 366	·161 245$^+$	·198 029
24	·020 635$^+$	·040 884	·060 759	·080 274	·099 439	·118 268	·136 770	·154 956	·190 421

Table B (*cont.*)

(2) Values of $\mathfrak{R}_2 = M_2 - M_1^2$

f_1 \ λ	0·5	1·0	1·5	2·0	2·5	3·0	3·5	4·0	5·0
2	−0·026 532	−0·090 643	−0·175 196	−0·269 055⁻	−0·365 139	−0·459 091	−0·548 369	−0·631 624	−0·778 254
3	− ·009 656	− ·033 727	− ·066 580	− ·104 320	− ·144 293	− ·184 718	− ·224 428	− ·262 689	− ·333 333
4	− ·004 606	− ·016 358	− ·032 807	− ·052 184	− ·073 221	− ·095 015⁺	− ·116 937	− ·138 550⁺	− ·179 798
5	−0·002 561	−0·009 215⁻	−0·018 710	−0·030 113	−0·042 726	−0·056 036	−0·069 664	−0·083 335⁻	−0·110 076
6	− ·001 573	− ·005 719	− ·011 728	− ·019 055⁻	− ·027 281	− ·036 086	− ·045 229	− ·054 526	− ·073 066
7	− ·001 037	− ·003 801	− ·007 857	− ·012 865⁻	− ·018 555⁻	− ·024 718	− ·031 189	− ·037 842	− ·051 319
8	− ·000 720	− ·002 657	− ·005 531	− ·009 114	− ·013 227	− ·017 725⁺	− ·022 493	− ·027 439	− ·037 590
9	− ·000 520	− ·001 932	− ·004 045⁻	− ·006 703	− ·009 779	− ·013 171	− ·016 795⁻	− ·020 583	− ·028 443
10	−0·000 388	−0·001 450⁻	−0·003 050⁺	−0·005 079	−0·007 443	−0·010 070	−0·012 893	−0·015 865⁻	−0·022 089
11	− ·000 298	− ·001 116	− ·002 358	− ·003 943	− ·005 802	− ·007 879	− ·010 126	− ·012 504	− ·017 525⁺
12	− ·000 233	− ·000 878	− ·001 862	− ·003 124	− ·004 614	− ·006 286	− ·008 105⁻	− ·010 039	− ·014 154
13	− ·000 186	− ·000 703	− ·001 496	− ·002 519	− ·003 731	− ·005 099	− ·006 593	− ·008 189	− ·011 607
14	− ·000 151	− ·000 572	− ·001 221	− ·002 061	− ·003 061	− ·004 195⁻	− ·005 438	− ·006 772	− ·009 643
15	−0·000 124	−0·000 472	−0·001 009	−0·001 708	−0·002 544	−0·003 494	−0·004 540	−0·005 666	−0·008 103
16	− ·000 103	− ·000 393	− ·000 844	− ·001 432	− ·002 137	−0·002 942	− ·003 831	− ·004 791	− ·006 878
17	− ·000 087	− ·000 332	− ·000 713	− ·001 213	− ·001 813	− ·002 501	− ·003 263	− ·004 088	− ·005 890
18	− ·000 074	− ·000 282	− ·000 608	− ·001 036	− ·001 552	− ·002 145⁻	− ·002 803	− ·003 518	− ·005 085⁻
19	− ·000 063	− ·000 242	− ·000 523	− ·000 892	− ·001 339	− ·001 853	− ·002 426	− ·003 049	− ·004 421
20	−0·000 055⁻	−0·000 209	−0·000 453	−0·000 774	−0·001 163	−0·001 612	−0·002 114	−0·002 661	−0·003 868
21	− ·000 047	− ·000 182	− ·000 395⁻	− ·000 676	− ·001 017	− ·001 412	− ·001 853	− ·002 336	− ·003 405⁺
22	− ·000 041	− ·000 160	− ·000 346	− ·000 593	− ·000 894	− ·001 243	− ·001 634	− ·002 062	− ·003 013
23	− ·000 036	− ·000 141	− ·000 305⁺	− ·000 524	− ·000 791	− ·001 101	− ·001 448	− ·001 830	− ·002 680
24	− ·000 032	− ·000 124	− ·000 271	− ·000 465⁺	− ·000 703	− ·000 979	− ·001 290	− ·001 632	− ·002 394

Table B. *The $\mathfrak{R}$-functions of equation* (9) (*cont.*)

(3) Values of $\mathfrak{R}_3 = M_3 - 3M_2M_1 + 2M_1^3$

f_1 \ λ	0·5	1·0	1·5	2·0	2·5	3·0	3·5	4·0	5·0
2	0·008 084	0·050 604	0·134 738	0·253 974	0·397 523	0·554 654	0·716 407	0·876 035$^+$	1·171 967
3	·001 696	·010 945$^-$	·029 997	·058 115$^+$	·093 357	·133 498	·176 474	·220 559	0·307 055$^-$
4	·000 535$^-$	·003 532	·009 897	·019 586	·032 106	·046 802	·063 007	·080 121	·115 158
5	0·000 213	0·001 431	0·004 082	0·008 217	0·013 689	0·020 266	0·027 688	0·035 706	0·052 675$^+$
6	·000 098	·000 673	·001 947	·003 974	·006 709	·010 060	·013 913	·018 152	·027 365$^+$
7	·000 051	·000 351	·001 029	·002 125$^-$	·003 628	·005 498	·007 683	·010 122	·015 543
8	·000 028	·000 198	·000 587	·001 225$^-$	·002 111	·003 229	·004 551	·006 047	·009 434
9	·000 017	·000 119	·000 356	·000 749	·001 301	·002 005$^+$	·002 848	·003 812	·006 028
10	0·000 011	0·000 075$^+$	0·000 226	0·000 479	0·000 839	0·001 303	0·001 862	0·002 508	0·004 014
11	·000 007	·000 049	·000 150$^-$	·000 319	·000 562	·000 877	·001 262	·001 709	·002 765$^-$
12	·000 005$^-$	·000 034	·000 102	·000 219	·000 388	·000 609	·000 881	·001 199	·001 959
13	·000 003	·000 023	·000 072	·000 155$^-$	·000 275$^+$	·000 435$^-$	·000 631	·000 863	·001 422
14	·000 002	·000 017	·000 052	·000 112	·000 200	·000 317	·000 462	·000 635$^-$	·001 054
15	0·000 002	0·000 012	0·000 038	0·000 083	0·000 148	0·000 236	0·000 345$^+$	0·000 476	0·000 796
16	·000 001	·000 009	·000 028	·000 062	·000 112	·000 178	·000 262	·000 363	·000 610
17	·000 001	·000 007	·000 022	·000 047	·000 086	·000 137	·000 202	·000 280	·000 475$^-$
18	·000 001	·000 005$^+$	·000 017	·000 037	·000 066	·000 107	·000 158	·000 220	·000 374
19	·000 000	·000 004	·000 013	·000 029	·000 052	·000 084	·000 125$^-$	·000 174	·000 298
20	—	0·000 003	0·000 010	0·000 023	0·000 042	0·000 067	0·000 100	0·000 140	0·000 240
21	—	·000 003	·000 008	·000 018	·000 033	·000 054	·000 081	·000 113	·000 195$^-$
22	—	·000 002	·000 007	·000 015$^-$	·000 027	·000 044	·000 066	·000 092	·000 160
23	—	·000 002	·000 005$^+$	·000 012	·000 022	·000 036	·000 054	·000 076	·000 132
24	—	·000 001	·000 004	·000 010	·000 018	·000 030	·000 045$^-$	·000 063	·000 110

Table B (*cont.*)

(4) Values of $\mathfrak{R}_4 = M_4 - 4M_3M_1 - 3M_2^2 + 12M_2M_1^2 - 6M_1^4$

f_1 \ λ	0·5	1·0	1·5	2·0	2·5	3·0	3·5	4·0	5·0
2	−0·003 823	−0·044 019	−0·162 067	−0·376 378	−0·682 037	−1·060 058	−1·486 109	−1·936 307	−2·832 241
3	− ·000 471	− ·005 640	− ·021 549	− ·051 831	− ·097 081	− ·155 660	−0·224 703	−0·300 922	−0·462 608
4	− ·000 100	− ·001 230	− ·004 835+	− ·011 951	− ·022 969	− ·037 743	− ·055 765−	−0·076 342	− ·122 208
5	−0·000 029	−0·000 363	−0·001 461	−0·003 690	−0·007 241	−0·012 138	−0·018 276	−0·025 476	−0·042 173
6	− ·000 010	− ·000 131	− ·000 536	− ·001 378	− ·002 751	− ·004 688	− ·007 171	− ·010 149	− ·017 281
7	− ·000 004	− ·000 054	− ·000 225+	− ·000 588	− ·001 192	− ·002 060	− ·003 194	− ·004 579	− ·007 988
8	− ·000 002	− ·000 025−	− ·000 105+	− ·000 278	− ·000 570	− ·000 997	− ·001 564	− ·002 268	− ·004 039
9	− ·000 001	− ·000 012	− ·000 053	− ·000 142	− ·000 295−	− ·000 521	− ·000 825−	− ·001 207	− ·002 190
10	—	−0·000 007	−0·000 029	−0·000 077	−0·000 162	−0·000 289	−0·000 461	−0·000 681	−0·001 256
11	—	− ·000 004	− ·000 016	− ·000 044	− ·000 094	− ·000 168	− ·000 271	− ·000 403	− ·000 754
12	—	− ·000 002	− ·000 010	− ·000 027	− ·000 056	− ·000 102	− ·000 166	− ·000 248	− ·000 470
13	—	− ·000 001	− ·000 006	− ·000 016	− ·000 035+	− ·000 064	− ·000 105−	− ·000 158	− ·000 303
14	—	− ·000 001	− ·000 004	− ·000 011	− ·000 023	− ·000 042	− ·000 068	− ·000 104	− ·000 201
15	—	−0·000 001	−0·000 002	−0·000 007	−0·000 015+	−0·000 024	−0·000 046	−0·000 070	−0·000 137
16	—	—	− ·000 002	− ·000 005−	− ·000 010	− ·000 019	− ·000 031	− ·000 048	− ·000 095−
17	—	—	− ·000 001	− ·000 003	− ·000 007	− ·000 013	− ·000 022	− ·000 034	− ·000 067
18	—	—	− ·000 001	− ·000 002	− ·000 005+	− ·000 009	− ·000 016	− ·000 024	− ·000 048
19	—	—	− ·000 001	− ·000 002	− ·000 004	− ·000 007	− ·000 011	− ·000 018	− ·000 035+
20	—	—	—	−0·000 001	−0·000 003	−0·000 005−	−0·000 008	−0·000 013	−0·000 026
21	—	—	—	− ·000 001	− ·000 002	− ·000 004	− ·000 006	− ·000 001	− ·000 020
22	—	—	—	− ·000 001	− ·000 001	− ·000 003	− ·000 005−	− ·000 007	− ·000 015+
23	—	—	—	—	− ·000 001	− ·000 002	− ·000 004	− ·000 005+	− ·000 012
24	—	—	—	—	− ·000 001	− ·000 002	− ·000 003	− ·000 004	− ·000 009

EDITORIAL NOTE

The preceding papers by Severo & Zelen and by Barton, David & O'Neill, both of which contain suggestions for approximating to the probability integral of non-central F, were received for publication within a few days of each other. The first paper uses the normal approximation to the cube root of both χ^2 and non-central χ^2, and then adopts the device (which seems first to have been suggested by David & Johnson, 1951) for determining isolated values of the probability integral of the ratio of two independent variables. The second paper first derives the cumulants of the logarithm of non-central F, and then applies to the distribution of this statistic, or rather to $z^* = \frac{1}{2}\log F^*$, various forms of inversion of an Edgeworth type expansion.

Severo & Zelen have compared their approximation with that of Patnaik which was based on representing the distribution of F^* by a central F distribution having the correct first two moments. It seems desirable also to compare their approximation with that of Barton *et al.*; this has been done by taking some of the cases examined by the latter authors in their Table 2 and applying to them the Severo & Zelen procedure.

The accompanying table suggests that at any rate for the combination of parameters chosen, there is on the whole little difference in accuracy between the Barton *et al.* Series II expansion and the Severo & Zelen method. It will be noted, however, that the results obtained by fitting Johnson's S_u distribution to that of z^*, using four moments, are almost invariably the most accurate. E.S.P.

REFERENCE

DAVID, F. N. & JOHNSON, N. L. (1951). *Biometrika*, **38**, 43–57.

Approximations to $P\{z^* < z_\alpha\} = P\{F^* < F_\alpha\}$

	Case: $\lambda = 0{\cdot}5, f_1 = f_2 = 10$				Case: $\lambda = 1{\cdot}0, f_1 = 10, f_2 = 20$			
		Error in approximation				Error in approximation		
α	Exact value	Series I	Series II	Severo & Zelen	Exact value	Series I	Series II	Severo & Zelen
0·005	0·9940	1	−1	−4	0·9918	−10	1	−1
·01	·9881	0	−2	−4	·9843	− 9	4	−2
·05	·9423	−3	1	−2	·9288	0	5	1
·50	·4699	7	−1	−1	·4321	9	−8	−4
·95	·0431	−4	0	2	·0364	2	5	2
·99	·0084	0	1	3	·0069	8	4	3
·995	·0041	1	1	3	·0034	7	0	3

Errors are given as units in the fourth decimal place.

	Case: $\lambda = 3{\cdot}0, \alpha = 0{\cdot}01$				Case: $\lambda = 3{\cdot}0, \alpha = 0{\cdot}05$			
		Error in approximations				Error in approximations		
	Exact value	Series II	S_u	Severo & Zelen	Exact value	Series II	S_u	Severo & Zelen
$f_1 = 2, f_2 = 2$	0·975	−4	0	−11	0·881	10	0	−5
4	·957	11	1	− 2	·824	4	−1	2
10	·920	10	1	1	·752	− 6	−3	7
20	·895	3	0	4	·718	−10	−6	10
	Case: $\lambda = 5{\cdot}0, \alpha = 0{\cdot}01$				Case: $\lambda = 5{\cdot}0, \alpha = 0{\cdot}05$			
$f_1 = 4, f_2 = 2$	0·978	−6	0	−12	0·892	4	1	−5
4	·960	0	0	− 2	·833	4	0	1
10	·916	4	1	1	·738	−1	0	3
20	·878	5	1	2	·681	−2	0	6

Errors are given as units in the third decimal place.

Biometrika (1960), **47**, 3 *and* 4, *p.* 433
Printed in Great Britain

Tables for making inferences about the variance of a normal distribution

BY D. V. LINDLEY, D. A. EAST AND P. A. HAMILTON
Statistical Laboratory, Cambridge

INTRODUCTION—BY D. V. LINDLEY

1. Neyman & Pearson (1936) showed that if we have a random sample of size n from a normal distribution of zero mean and unknown variance σ^2, then an unbiased region of type A with size α for testing the null hypothesis that $\sigma = \sigma_0$ against the alternative $\sigma \neq \sigma_0$ is obtained by using the statistic

$$v = \sum_{i=1}^{n} x_i^2/\sigma_0^2 \tag{1}$$

(which, on the null hypothesis, has a χ^2-distribution on n degrees of freedom) and rejecting whenever $v \leqslant v_1$ or $v \geqslant v_2$, where v_1 and v_2 are so chosen that

$$p(v \leqslant v_1|\sigma_0) + p(v \geqslant v_2|\sigma_0) = \alpha \tag{2}$$

and

$$v_1^{\frac{1}{2}n} e^{-\frac{1}{2}v_1} = v_2^{\frac{1}{2}n} e^{-\frac{1}{2}v_2}. \tag{3}$$

(Cf. equations (138)–(142) of the paper cited.) By using the results of Neyman (1935) it is possible to show that if the mean μ is also unknown, and therefore a nuisance parameter, an unbiased region of type B with size α for testing the same hypothesis is obtained by using the statistic

$$v' = \sum_{i=1}^{n} (x_i - \bar{x})^2/\sigma_0^2$$

and rejecting whenever $v' \leqslant v_1'$ or $v' \geqslant v_2'$, where v_1' and v_2' satisfy (2), for all μ, and (3), with primes attached to the v's and n replaced by $n-1$. In both these situations then a two-sided test based on the χ^2-distribution is used but instead of using equal tail-areas (with $\frac{1}{2}\alpha$ in each tail) the areas are chosen so that the critical values obey (3). Notice that, introducing the same constant $\{2^{\frac{1}{2}(n+2)}\Gamma[\frac{1}{2}(n+2)]\}^{-1}$ on both sides of (3), the constraint (3) says that the ordinates of a χ^2-distribution on $(n+2)$ degrees of freedom must be equal at the critical values v_1 and v_2. The usual connexions between tests of hypotheses and confidence intervals enable us to say that the interval for σ^2 derived in the usual way from this test, i.e. from

$$p(v_1 \leqslant (\Sigma x^2)/\sigma^2 \leqslant v_2) = 1 - \alpha$$

is the best unbiased confidence interval. The present tables give values of $v_1 = \chi_1^2$ and $v_2 = \chi_2^2$ for degrees of freedom proceeding by integers from 1 to 100, and $\alpha = 0{\cdot}05$, $0{\cdot}01$ and $0{\cdot}001$.

The tables were not prepared in order to derive the best unbiased tests but for a different reason. Let us consider again a random sample of size n from a normal distribution of zero mean and unknown variance and consider what we can say about the variance on the basis of the sample. If we use a Bayes approach there are strong, but not completely convincing, reasons for assuming *a priori* that $\ln \sigma$ is uniformly distributed over the whole

real line. It is then easy to see that the posterior distribution of σ^2 is such that $\left(\sum_{i=1}^{n} x_i^2\right)\Big/\sigma^2$ is distributed as χ^2 on n d.f. (see, for example, Jeffreys (1948), § 3·41). That is

$$p(\sigma^2|x_1, \ldots, x_n) \propto \frac{1}{\sigma^{n+2}} \exp\left[-(\Sigma x^2)/2\sigma^2\right]. \tag{4}$$

This posterior distribution expresses our knowledge of the variance after the sample has been taken. It would be enough to quote this distribution in any given case but it is obviously convenient, and in line with modern practice, to quote an interval within which σ^2 has high (e.g. 95 %) posterior probability of lying. Furthermore, this interval should be such that points within the interval have higher posterior density than points without it, and this additional requirement would determine a unique interval, which would be the shortest containing a given amount of the posterior probability. Unfortunately, the requirement is not invariant under a transformation of the parameter, because a density is not invariant, and we have to specify the function of σ that interests us. Since we have assumed $\ln \sigma$ uniform *a priori* it seems natural to look at the same function *a posteriori*. The posterior distribution of $u = \ln \sigma$ is proportional to

$$e^{-nu} \exp\left[-(\Sigma x^2)/2e^{2u}\right]$$

and our interval of u contains points where this exceeds some value, which itself is chosen so that the interval contains a predetermined posterior probability. Hence the critical values u_1 and u_2 satisfy

$$e^{-nu_1} \exp\left[-(\Sigma x^2)/2e^{2u_1}\right] = e^{-nu_2} \exp\left[-(\Sigma x^2)/2e^{2u_2}\right],$$

or letting

$$v = (\Sigma x^2)/\sigma^2 = (\Sigma x^2)/e^{2u},$$

$$v_1^{\frac{1}{2}n} e^{-\frac{1}{2}v_1} = v_2^{\frac{1}{2}n} e^{-\frac{1}{2}v_2}$$

which is (3) again. Since v has a χ^2-distribution even when σ^2 (and not Σx^2) is regarded as the variable the problem of finding the shortest interval in $\ln \sigma$ containing an amount $(1-\alpha)$ of the posterior distribution is computationally the same as the problem of finding the best unbiased confidence interval. The parallelism persists if μ is unknown and assumed to have a uniform prior distribution independent of $\ln \sigma$: the degrees of freedom are reduced by one in the usual way. The reason for the parallelism is that the distribution of $\frac{1}{2} \ln (\Sigma x^2)$ has $\ln \sigma$ as a location parameter and hence there is a symmetry between operations in the sample space and parameter space when the latter is described by $\ln \sigma$.

The tables also provide solutions of the following problem: find the shortest interval of the posterior distribution of σ^2 containing probability $(1-\alpha)$. It is only necessary to decrease the degrees of freedom by 2 and use the tabulated values. The proof of this is straightforward on integrating by parts. With one and two degrees of freedom the mode of the χ^2-distribution is at the origin and the shortest interval is one-sided and may be found from the usual tables.

2. The tables were computed by Method II of Neyman & Pearson (1936, Appendix II) with one modification. We require to find an interval (χ_1^2, χ_2^2) of the χ^2-distribution on ν degrees of freedom containing probability $(1-\alpha)$ (equation (2)) which is such that the ordinates of the χ^2-distribution on $(\nu+2)$ degrees of freedom at χ_1^2 and χ_2^2 are equal (equation (3)). The values are clearly unique. For any ν we have an approximation $\overline{\chi}_1^2$ to the required

value of $\overline{\chi}_1^2$ obtained from the lower $\frac{1}{2}\alpha$ point of the distribution. With this value we obtain a solution $\overline{\chi}_2^2$ of the equation in x

$$(\overline{\chi}_1^2)^{\frac{1}{2}n} \exp(-\tfrac{1}{2}\overline{\chi}_1^2) = x^{\frac{1}{2}n} e^{-\frac{1}{2}x} \tag{5}$$

derived from (3). With these values of $\overline{\chi}_1^2$ and $\overline{\chi}_2^2$ the probability $1-\alpha$ will be too small and we repeat the process with a smaller value of χ_1^2, continuing until we converge to the required probability. The only way we differ from Neyman & Pearson is in the solution of (5). The method proposed by them is to start from a guess $\hat{\chi}_2^2$ (for example, the upper $\frac{1}{2}\alpha$ point of χ^2) and use an iterative process which is of first order (see Hartree (1952), § 9·3). Instead we used Newton's iterative procedure (ibid. § 9·31) which is of second order and therefore converges more rapidly to the required value. For even ν the integrals of χ^2 involve only the exponential function. The calculations for these values were carried out on EDSAC II at the Mathematical Laboratory, Cambridge, with a programme drawn up by two of us (D. A. E. and P. A. H.). This made use of the values of the exponential function which are held in the reserved store of the machine as a permanent subroutine. They were carried through to seven significant figures and rounded off to the five given in the tables. Differences were checked by hand and used to interpolate approximate values for small odd degrees of freedom, other than one. These interpolated values were used as starting values for the same iterative procedure carried through on a desk machine using tables of the normal distribution function. (This function is not available as a permanent subroutine.) The procedure was carried through with each odd degree of freedom until the interpolate using fourth differences was found to be accurate to five significant figures: the values for higher odd degrees of freedom were then obtained by interpolation. The values for $\nu = 1$ were found similarly.

Since these calculations were carried out Tate & Klett (1959) have published a similar table (Table 680). We have checked our values against theirs and have found discrepancies, particularly in the 0·1 % values. We believe, after some further checking, that our values are correct and Klett writes in a private communication that he thinks some error may have been made in their calculations. They also mention some tables of Rao's (1952, p. 223): these also seem to be in error.

We are grateful to the Director of the Mathematical Laboratory for permission to use the machine, and to the staff of the Laboratory for their co-operation.

References

Hartree, D. R. (1952). *Numerical Analysis.* Oxford: Clarendon Press.

Jeffreys, H. (1948). *Theory of Probability.* Oxford: Clarendon Press. Second edition.

Neyman, J. (1935). Sur la vérification des hypothèses statistiques composées. *Bull. Soc. Math. Fr.* **63**, 1–21.

Neyman, J. & Pearson, E. S. (1936). Contributions to the theory of testing statistical hypotheses. I. *Statist. Res. Mem.* **1**, 1–37.

Rao, C. R. (1952). *Advanced Statistical Methods in Biometric Research.* New York: John Wiley and Sons, Inc.

Tate, R. F. & Klett, G. W. (1959). Optimal confidence intervals for the variance of a normal distribution. *J. Amer. Statist. Ass.* **54**, 674–82.

Two-sided tests for the χ^2-distribution

	5%		1%		0·1%	
ν	χ^2_1	χ^2_2	χ^2_1	χ^2_2	χ^2_1	χ^2_2
1	$0{\cdot}0^3 31593$	7·8168	$0{\cdot}0^3 13422$	11·345	$0{\cdot}0^5 14026$	16·266
2	0·084727	9·5303	0·017469	13·285	$0{\cdot}0^2 18055$	18·468
3	0·29624	11·191	0·101048	15·127	0·022097	20·524
4	0·60700	12·802	0·26396	16·901	0·083097	22·486
5	0·98923	14·369	0·49623	18·621	0·19336	24·378
6	1·4250	15·897	0·78565	20·296	0·35203	26·214
7	1·9026	17·392	1·1221	21·931	0·55491	28·004
8	2·4139	18·860	1·4978	23·533	0·79722	29·754
9	2·9532	20·305	1·9069	25·106	1·0745	31·469
10	3·5162	21·729	2·3444	26·653	1·3827	33·154
11	4·0995	23·135	2·8069	28·178	1·7185	34·812
12	4·7005	24·525	3·2912	29·683	2·0791	36·446
13	5·3171	25·900	3·7949	31·170	2·4620	38·058
14	5·9477	27·263	4·3161	32·641	2·8651	39·650
15	6·5908	28·614	4·8530	34·097	3·2865	41·225
16	7·2453	29·955	5·4041	35·540	3·7248	42·783
17	7·9100	31·285	5·9683	36·971	4·1786	44·325
18	8·5842	32·607	6·5444	38·390	4·6468	45·854
19	9·2670	33·921	7·1316	39·798	5·1281	47·370
20	9·9579	35·227	7·7289	41·197	5·6218	48·874
21	10·656	36·525	8·3358	42·586	6·1269	50·366
22	11·361	37·818	8·9515	43·967	6·6428	51·848
23	12·073	39·103	9·5755	45·340	7·1688	53·320
24	12·791	40·383	10·2073	46·706	7·7043	54·782
25	13·514	41·658	10·846	48·064	8·2487	56·236
26	14·243	42·927	11·492	49·416	8·8016	57·682
27	14·977	44·192	12·145	50·761	9·3625	59·119
28	15·716	45·451	12·803	52·100	9·9310	60·549
29	16·459	46·707	13·468	53·434	10·507	61·972
30	17·206	47·958	14·138	54·762	11·089	63·388
31	17·958	49·205	14·813	56·085	11·678	64·798
32	18·713	50·448	15·494	57·403	12·274	66·202
33	19·472	51·688	16·179	58·716	12·875	67·599
34	20·235	52·924	16·869	60·025	13·482	68·991
35	21·001	54·157	17·563	61·330	14·094	70·378
36	21·771	55·386	18·261	62·630	14·712	71·759
37	22·543	56·613	18·964	63·927	15·335	73·136
38	23·319	57·836	19·670	65·219	15·963	74·507
39	24·097	59·057	20·380	66·508	16·595	75·874
40	24·879	60·275	21·094	67·793	17·232	77·236
41	25·663	61·490	21·811	69·075	17·873	78·595
42	26·449	62·703	22·531	70·354	18·518	79·948
43	27·238	63·913	23·255	71·629	19·168	81·298
44	28·029	65·121	23·982	72·901	19·821	82·645
45	28·823	66·327	24·712	74·170	20·478	83·987
46	29·619	67·530	25·445	75·437	21·139	85·326
47	30·417	68·731	26·181	76·700	21·803	86·661
48	31·218	69·931	26·919	77·961	22·471	87·992
49	32·020	71·128	27·660	79·220	23·142	89·321
50	32·824	72·323	28·404	80·475	23·816	90·646

	5%		1%		0·1%	
ν	χ_1^2	χ_2^2	χ_1^2	χ_2^2	χ_1^2	χ_2^2
51	33·630	73·516	29·150	81·729	24·494	91·968
52	34·439	74·708	29·898	82·979	25·174	93·287
53	35·248	75·897	30·649	84·228	25·858	94·603
54	36·060	77·085	31·403	85·474	26·544	95·916
55	36·873	78·271	32·158	86·718	27·233	97·227
56	37·689	79·456	32·916	87·960	27·925	98·535
57	38·505	80·639	33·675	89·200	28·620	99·840
58	39·323	81·820	34·437	90·437	29·317	101·142
59	40·143	83·000	35·201	91·673	30·016	102·442
60	40·965	84·178	35·967	92·907	30·719	103·74
61	41·787	85·355	36·735	94·139	31·423	105·03
62	42·612	86·531	37·504	95·369	32·130	106·33
63	43·437	87·705	38·276	96·597	32·839	107·62
64	44·264	88·878	39·049	97·823	33·551	108·91
65	45·092	90·049	39·824	99·048	34·264	110·19
66	45·922	91·219	40·600	100·271	34·980	111·48
67	46·753	92·388	41·379	101·492	35·698	112·76
68	47·585	93·555	42·159	102·71	36·418	114·04
69	48·418	94·722	42·940	103·93	37·140	115·32
70	49·253	95·887	43·723	105·15	37·864	116·59
71	50·089	97·051	44·508	106·36	38·590	117·87
72	50·926	98·214	45·294	107·58	39·317	119·14
73	51·764	99·376	46·081	108·79	40·047	120·41
74	52·603	100·536	46·870	110·00	40·778	121·68
75	53·443	101·696	47·661	111·21	41·511	122·94
76	54·284	102·85	48·452	112·42	42·246	124·21
77	55·126	104·01	49·245	113·62	42·983	125·47
78	55·969	105·17	50·040	114·83	43·721	126·73
79	56·814	106·32	50·836	116·03	44·461	127·99
80	57·659	107·48	51·633	117·23	45·203	129·25
81	58·505	108·63	52·431	118·44	45·946	130·51
82	59·352	109·79	53·230	119·64	46·690	131·76
83	60·200	110·94	54·031	120·84	47·436	133·02
84	61·049	112·09	54·833	122·03	48·184	134·27
85	61·899	113·24	55·636	123·23	48·933	135·52
86	62·750	114·39	56·440	124·43	49·684	136·77
87	63·601	115·54	57·245	125·62	50·436	138·02
88	64·454	116·68	58·052	126·81	51·189	139·26
89	65·307	117·83	58·859	128·01	51·944	140·51
90	66·161	118·98	59·668	129·20	52·700	141·75
91	67·016	120·12	60·477	130·39	53·457	142·99
92	67·871	121·26	61·288	131·58	54·216	144·23
93	68·728	122·41	62·100	132·76	54·976	145·47
94	69·585	123·55	62·912	133·95	55·738	146·71
95	70·443	124·69	63·726	135·14	56·500	147·95
96	71·302	125·83	64·540	136·32	57·264	149·19
97	72·161	126·97	65·356	137·51	58·029	150·42
98	73·021	128·11	66·172	138·69	58·795	151·66
99	73·882	129·25	66·990	139·87	59·562	152·89
100	74·744	130·39	67·808	141·05	60·331	154·12

The values of χ_1^2 and χ_2^2 satisfy the equations

$$\int_0^{\chi_1^2} + \int_{\chi_2^2}^{\infty} 2^{-\frac{1}{2}\nu}\{\Gamma(\tfrac{1}{2}\nu)\}^{-1} e^{-\frac{1}{2}x}\, x^{\frac{1}{2}\nu-1}\, dx = \alpha$$

and

$$(\chi_1^2)^{\frac{1}{2}\nu}\, e^{-\frac{1}{2}\chi_1^2} = (\chi_2^2)^{\frac{1}{2}\nu}\, e^{-\frac{1}{2}\chi_2^2}$$

for $\alpha = 0{\cdot}05$, $0{\cdot}01$ and $0{\cdot}001$.

Biometrika (1960), **47**, 3 *and* 4, *p.* 439
Printed in Great Britain

Tables for the solution of the exponential equation, $\exp(-a)+ka=1$

BY D. E. BARTON, F. N. DAVID† AND M. MERRINGTON
University College London

David & Johnson (1952) pointed out that the estimation by maximum likelihood of the parameter of a Poisson distribution when the zero group is missing reduces to the solution of the equation

$$\bar{x} = \frac{\hat{\lambda}}{1-e^{-\hat{\lambda}}},$$

where $\bar{x}$ is the mean of the N known observations and λ is the Poisson parameter.

Recently (Barton & David, 1961) we have been investigating approximations to the difference quotients of zero, i.e. to

$$\frac{\Delta^n 0^i}{n!}$$

for values of n outside the range of existing tables. Various approximations have recently been proposed, Arfwedson (1951), Binet & Szekeres (1957) and Hsu (1948). Incomparably the best of any put forward are those given by Laplace, *Théorie Analytique des Probabilités* (1820), Livre 1, p. 162. Laplace's approximation to $\Delta^n 0^i$ depends on the solution for a of

$$\frac{i+1}{a} = \frac{ne^a}{e^a-1}.$$

These two requirements for the solution of this exponential equation will possibly not be the only contexts in which it intervenes in statistical practice. From the combinatorial point of view they seem to us, by themselves, to be sufficient justification for the computation of the table which is given in this paper.

CONSTRUCTION OF THE TABLE

We considered the solution of the equation

$$e^{-a}+ka = 1,$$

where k is known, and the solution for a is required. For values of $k = 0{\cdot}050\ (0{\cdot}001)\ 0{\cdot}190$, a was found approximately by building up a value of $k^{-1}-a$ from third differences, choosing an approximate value for a and solving for a by iteration. For $k = 0{\cdot}190(0{\cdot}001)1{\cdot}000$ a table of the approximate value of a was built up assuming constant fifth differences, the exact value of a then being found by iteration. For $k < 0{\cdot}05$ it is enough to assume

$$a = \frac{1}{k},$$

since this approximation overestimates a by less than $4{\cdot}2 \times 10^{-8}$. The table was checked on the Mercury computer with the help of the staff of the University of London Computer Unit.

† With the partial support of the office of Naval Research while at the University of California, Berkeley.

Example 1

Van Rest (1937) in his paper on *Examples of Statistical Methods in Forest Products Research* gives the following table in which the expected frequencies (i) are calculated by treating $x-1$ as a Poisson variable, where x is the number of vessels in a group.

Frequency of occurrence of different-sized groups of vessels in the wood of Shorea leprosula

No. of vessels in group	Frequency	Expected frequency (i)	Expected frequency (ii)
1	843	838·5	843
2	143	151·9	144
3	18	13·7	17
4	1	0·8	1
Total	1005	1005	1005

The mean of the distribution is $\bar{x} = 1{\cdot}1811$ whence $k = 1/\bar{x} = 0{\cdot}8467$. From the table, $a = \hat{\lambda} = 0{\cdot}3426$. Poisson probabilities with this parameter were calculated. The expected frequencies (ii) are these probabilities divided by their total of 0·2901 and multiplied by 1005. The fit by maximum likelihood, assuming that the distribution is a Poisson with zero group missing, is seen to give better results than the method of graduation used by Van Rest.

Example 2

Stevens (1937) gave a table of $\dfrac{\Delta^n 0^i}{n!}$

for $n = 2(1)\ldots i$, $i = 2(1)25$. The values of the difference quotients of zero are accordingly only required outside this range and for test purposes, calculations at $i = 20$ will provide a severe check on any approximation. Laplace gave what is equivalent to the relation

$$\frac{\Delta^n 0^i}{n!} = \frac{i^2\Gamma(i)}{n!\,a^{i+1}} \frac{(e^a-1)^n}{\sqrt{(2\pi)\left[\dfrac{i+1}{a^2} - \dfrac{ne^a}{(e^a-1)^2}\right]^{\frac{1}{2}}}} \left\{1 + \frac{15l'^2 - 12ll''}{16l^3} + \ldots\right\},$$

where a is the solution of

$$\frac{i+1}{a} = \frac{ne^a}{e^a-1}$$

and

$$l = -\frac{i+1}{2a^2} - \frac{n}{2}\left(\frac{e^a}{e^a-1}\right) + \frac{n}{2}\left(\frac{e^a}{e^a-1}\right)^2,$$

$$l' = -\frac{i+1}{3a^3} + \frac{n}{6}\left(\frac{e^a}{e^a-1}\right) - \frac{n}{2}\left(\frac{e^a}{e^a-1}\right)^2 + \frac{n}{3}\left(\frac{e^a}{e^a-1}\right)^3,$$

$$l'' = -\frac{i+1}{4a^4} - \frac{n}{24}\left(\frac{e^a}{e^a-1}\right) + \frac{7n}{24}\left(\frac{e^a}{e^a-1}\right)^2 - \frac{n}{2}\left(\frac{e^a}{e^a-1}\right)^3 + \frac{n}{4}\left(\frac{e^a}{e^a-1}\right)^4.$$

For $i = 20$ the approximate values of the difference quotients were calculated from this formula, from De Moivre's (1756) formula, and from formulae recently proposed by Binet & Szekeres (1957), Arfwedson (1951), and Hsu (1948) who had been anticipated by Cayley (1881). The proportionate errors, namely (approximate value–true value)/(true value), multiplied by 100 are given in the following table.

Relative error (approximate value–true value)/(true value) expressed as a percentage using five different approximations for the difference quotients of zero ($i = 20$)

n	Laplace	De Moivre	Binet & Szekeres	Arfwedson	Hsu
3	0·005	0·000	0·133	2·947	−100·000
4	0·012	0·006	0·266	2·407	−100·000
5	0·008	0·101	0·405	2·093	−99·987
6	−0·001	0·633	1·036	1·893	−99·823
7	−0·010	2·309	3·167	1·741	−98·755
8	−0·019	6·190	8·302	1·631	−94·694
9	−0·026	13·780	18·636	1·557	−84·660
10	−0·032	27·389	37·757	1·514	−67·397
11	−0·037	51·000	72·343	1·504	−45·775
12	−0·042	92·238	136·294	1·530	−25·538
13	−0·046	167·151	261·639	1·598	−11·256
14	−0·049	313·045	530·772	1·726	−3·721
15	−0·053	627·365	1185·312	1·940	−0·848

Laplace's approximation becomes closer to the true value with increasing i. It is remarkably good and can obviously be used with confidence for the calculation of any value or set of values outside the range of Stevens's tables.

It is of interest to note that, if we write

$$x_1 = -\frac{1}{k}, \quad y = x_1 e^{x_1}$$

then the second real root of $xe^x = y,$

is $$x_2 = a - \frac{1}{k}.$$

Thus our table may be considered, essentially, as tabulating, for negative y, that of the two roots which is less in modulus than unity, in terms of the other. When y is positive the equation has only one real root (necessary, incidentally, for the application of the Binet–Szekeres method). We shall treat the equation in this range, in a future paper, since it is a useful auxiliary in Laplace's expansion for the Differences of the Reciprocals of Unity.

References

Arfwedson, G. (1951). A probability distribution connected with Stirling's second class numbers. *Skand. Aktuar. Tidskr.* **34**, 121–32.

Barton, D. E. & David, F. N. (1961). *Combinatorial Chance.* London: Charles Griffin and Co. (in the Press).

Binet, F. E. & Szekeres, G. (1957). On Borel fields over finite sets. *Ann. Math. Statist.* **28**, 494–8.

Cayley, A. (1881). Table of $\Delta^m O^n \div \Pi(m)$ up to $m = n = 20$. *Trans. Camb. Phil. Soc.* **13**, 1–4.

David, F. N. & Johnson, N. L. (1952). The truncated Poisson. *Biometrics*, **8**, 275–85.

De Moivre, A. (1756). *Doctrine of Chances*, 3rd edition.

Hsu, L. C. (1948). Note on an asymptotic expansion of the nth difference of zero. *Ann. Math. Statist.* **19**, 273–7.

Laplace, P. S. (1820). *Théorie Analytique des Probabilités.* Paris: Courcier.

Van Rest, E. D. (1937). Examples of statistical methods in forest products research. *J.R. Statist. Soc.* Suppl. **4**, 184–203.

Stevens, W. L. (1937). Significance of grouping. *Ann. Eugen., Lond.*, **8**, 57–69.

Table for the solution of the equation $\exp(-a) + ka = 1$

k	a	k	a	k	a	k	a	k	a
0·050	20·0000000	0·100	9·9995458	0·150	6·6581095	0·200	4·9651142	0·250	3·9206904
·051	19·6078431	·101	9·9004937	·151	6·6136295	·201	4·9395120	·251	3·9037180
·052	19·2307691	·102	9·8033798	·152	6·5697225	·202	4·9141487	·252	3·8868671
·053	18·8679244	·103	9·7081477	·153	6·5263771	·203	4·8890207	·253	3·8701363
·054	18·5185184	·104	9·6147429	·154	6·4835823	·204	4·8641246	·254	3·8535241
0·055	18·1818180	0·105	9·5231129	0·155	6·4413272	0·205	4·8394568	0·255	3·8370292
·056	17·8571425	·106	9·4332073	·156	6·3996012	·206	4·8150140	·256	3·8206502
·057	17·5438592	·107	9·3449775	·157	6·3583940	·207	4·7907929	·257	3·8043857
·058	17·2413787	·108	9·2583768	·158	6·3176956	·208	4·7667902	·258	3·7882344
·059	16·9491518	·109	9·1733599	·159	6·2774962	·209	4·7430026	·259	3·7721949
0·060	16·6666657	0·110	9·0898836	0·160	6·2377864	0·210	4·7194272	0·260	3·7562660
·061	16·3934414	·111	9·0079060	·161	6·1985568	·211	4·6960608	·261	3·7404463
·062	16·1290307	·112	8·9273866	·162	6·1597983	·212	4·6729003	·262	3·7247347
·063	15·8730138	·113	8·8482865	·163	6·1215021	·213	4·6499430	·263	3·7091299
·064	15·6249974	·114	8·7705681	·164	6·0836596	·214	4·6271857	·264	3·6936307
0·065	15·3846122	0·115	8·6941952	0·165	6·0462625	0·215	4·6046258	0·265	3·6782358
·066	15·1515112	·116	8·6191326	·166	6·0093024	·216	4·5822605	·266	3·6629441
·067	14·9253682	·117	8·5453466	·167	5·9727714	·217	4·5600869	·267	3·6477545
·068	14·7058763	·118	8·4728044	·168	5·9366617	·218	4·5381024	·268	3·6326657
·069	14·4927463	·119	8·4014745	·169	5·9009657	·219	4·5163044	·269	3·6176767
0·070	14·2857054	0·120	8·3313262	0·170	5·8656758	0·220	4·4946903	0·270	3·6027863
·071	14·0844963	·121	8·2623301	·171	5·8307850	·221	4·4732575	·271	3·5879935
·072	13·8888760	·122	8·1944575	·172	5·7962859	·222	4·4520036	·272	3·5732972
·073	13·6986147	·123	8·1276809	·173	5·7621718	·223	4·4309262	·273	3·5586962
·074	13·5134952	·124	8·0619734	·174	5·7284359	·224	4·4100228	·274	3·5441897
0·075	13·3333117	0·125	7·9973091	0·175	5·6950714	0·225	4·3892910	0·275	3·5297765
·076	13·1578693	·126	7·9336629	·176	5·6620721	·226	4·3687286	·276	3·5154556
·077	12·9869832	·127	7·8710106	·177	5·6294315	·227	4·3483333	·277	3·5012261
·078	12·8204781	·128	7·8093287	·178	5·5971436	·228	4·3281028	·278	3·4870869
·079	12·6581876	·129	7·7485942	·179	5·5652022	·229	4·3080351	·279	3·4730370
0·080	12·4999534	0·130	7·6887851	0·180	5·5336015	0·230	4·2881278	0·280	3·4590756
·081	12·3456253	·131	7·6298800	·181	5·5023357	·231	4·2683789	·281	3·4452017
·082	12·1950603	·132	7·5718581	·182	5·4713992	·232	4·2487864	·282	3·4314143
·083	12·0481222	·133	7·5146991	·183	5·4407865	·233	4·2293482	·283	3·4177125
·084	11·9046814	·134	7·4583837	·184	5·4104922	·234	4·2100622	·284	3·4040955
0·085	11·7646144	0·135	7·4028927	0·185	5·3805110	0·235	4·1909266	0·285	3·3905624
·086	11·6278033	·136	7·3482078	·186	5·3508378	·236	4·1719393	·286	3·3771122
·087	11·4941358	·137	7·2943110	·187	5·3214675	·237	4·1530985	·287	3·3637441
·088	11·3635044	·138	7·2411851	·188	5·2923952	·238	4·1344024	·288	3·3504572
·089	11·2358068	·139	7·1888131	·189	5·2636161	·239	4·1158490	·289	3·3372508
0·090	11·1109450	0·140	7·1371786	0·190	5·2351255	0·240	4·0974366	0·290	3·3241240
·091	10·9888254	·141	7·0862658	·191	5·2069187	·241	4·0791634	·291	3·3110759
·092	10·8693583	·142	7·0360593	·192	5·1789912	·242	4·0610277	·292	3·2981058
·093	10·7524581	·143	6·9865438	·193	5·1513386	·243	4·0430277	·293	3·2852129
·094	10·6380427	·144	6·9377049	·194	5·1239566	·244	4·0251617	·294	3·2723964
0·095	10·5260334	0·145	6·8895283	0·195	5·0968408	0·245	4·0074282	0·295	3·2596555
·096	10·4163548	·146	6·8420002	·196	5·0699872	·246	3·9898254	·296	3·2469895
·097	10·3089347	·147	6·7951072	·197	5·0433917	·247	3·9723518	·297	3·2343976
·098	10·2037037	·148	6·7488362	·198	5·0170503	·248	3·9550058	·298	3·2218790
·099	10·1005954	·149	6·7031744	·199	4·9909591	·249	3·9377858	·299	3·2094331
0·100	9·9995458	0·150	6·6581095	0·200	4·9651142	0·250	3·9206904	0·300	3·1970591

N.B. for $k < 0{\cdot}05$, take $a = 1/k$.

k	a	k	a	k	a	k	a	k	a
0·300	3·1970591	0·350	2·6566127	0·400	2·2316119	0·450	1·8847348	0·500	1·5936243
·301	3·1847564	·351	2·6471411	·401	2·2240016	·451	1·8784240	·501	1·5882633
·302	3·1725241	·352	2·6377143	·402	2·2164218	·452	1·8721351	·502	1·5829188
·303	3·1603617	·353	2·6283321	·403	2·2088724	·453	1·8658679	·503	1·5775904
·304	3·1482684	·354	2·6189939	·404	2·2013532	·454	1·8596224	·504	1·5722783
0·305	3·1362436	0·355	2·6096996	0·405	2·1938638	0·455	1·8533984	0·505	1·5669823
·306	3·1242866	·356	2·6004487	·406	2·1864042	·456	1·8471958	·506	1·5617022
·307	3·1123968	·357	2·5912409	·407	2·1789740	·457	1·8410145	·507	1·5564381
·308	3·1005735	·358	2·5820758	·408	2·1715732	·458	1·8348542	·508	1·5511898
·309	3·0888160	·359	2·5729530	·409	2·1642015	·459	1·8287149	·509	1·5459572
0·310	3·0771238	0·360	2·5638723	0·410	2·1568586	0·460	1·8225965	0·510	1·5407403
·311	3·0654961	·361	2·5548333	·411	2·1495444	·461	1·8164989	·511	1·5355389
·312	3·0539325	·362	2·5458356	·412	2·1422588	·462	1·8104218	·512	1·5303531
·313	3·0424323	·363	2·5368790	·413	2·1350014	·463	1·8043652	·513	1·5251826
·314	3·0309949	·364	2·5279631	·414	2·1277721	·464	1·7983290	·514	1·5200275
0·315	3·0196198	0·365	2·5190875	0·415	2·1205707	0·465	1·7923131	0·515	1·5148876
·316	3·0083062	·366	2·5102520	·416	2·1133971	·466	1·7863172	·516	1·5097628
·317	2·9970537	·367	2·5014563	·417	2·1062510	·467	1·7803413	·517	1·5046532
·318	2·9858617	·368	2·4927000	·418	2·0991322	·468	1·7743854	·518	1·4995585
·319	2·9747297	·369	2·4839828	·419	2·0920406	·469	1·7684491	·519	1·4944788
0·320	2·9636570	0·370	2·4753044	0·420	2·0849759	0·470	1·7625325	0·520	1·4894139
·321	2·9526432	·371	2·4666645	·421	2·0779381	·471	1·7566355	·521	1·4843637
·322	2·9416876	·372	2·4580629	·422	2·0709268	·472	1·7507578	·522	1·4793282
·323	2·9307898	·373	2·4494991	·423	2·0639420	·473	1·7448995	·523	1·4743074
·324	2·9199493	·374	2·4409730	·424	2·0569834	·474	1·7390603	·524	1·4693011
0·325	2·9091655	0·375	2·4324843	0·425	2·0500510	0·475	1·7332402	0·525	1·4643092
·326	2·8984379	·376	2·4240326	·426	2·0431444	·476	1·7274391	·526	1·4593317
·327	2·8877660	·377	2·4156176	·427	2·0362635	·477	1·7216568	·527	1·4543685
·328	2·8771494	·378	2·4072391	·428	2·0294083	·478	1·7158932	·528	1·4494195
·329	2·8665874	·379	2·3988969	·429	2·0225784	·479	1·7101483	·529	1·4444847
0·330	2·8560797	0·380	2·3905906	0·430	2·0157738	0·480	1·7044219	0·530	1·4395640
·331	2·8456257	·381	2·3823199	·431	2·0089942	·481	1·6987139	·531	1·4346573
·332	2·8352249	·382	2·3740846	·432	2·0022395	·482	1·6930242	·532	1·4297645
·333	2·8248770	·383	2·3658845	·433	1·9955095	·483	1·6873527	·533	1·4248856
·334	2·8145814	·384	2·3577192	·434	1·9888042	·484	1·6816993	·534	1·4200205
0·335	2·8043377	0·385	2·3495885	0·435	1·9821232	0·485	1·6760639	0·535	1·4151691
·336	2·7941455	·386	2·3414922	·436	1·9754665	·486	1·6704464	·536	1·4103313
·337	2·7840042	·387	2·3334300	·437	1·9688339	·487	1·6648467	·537	1·4055072
·338	2·7739134	·388	2·3254016	·438	1·9622253	·488	1·6592647	·538	1·4006965
·339	2·7638728	·389	2·3174068	·439	1·9556405	·489	1·6537002	·539	1·3958994
0·340	2·7538818	0·390	2·3094453	0·440	1·9490792	0·490	1·6481533	0·540	1·3911155
·341	2·7439401	·391	2·3015170	·441	1·9425415	·491	1·6426237	·541	1·3863450
·342	2·7340472	·392	2·2936214	·442	1·9360271	·492	1·6371114	·542	1·3815878
·343	2·7242027	·393	2·2857585	·443	1·9295360	·493	1·6316164	·543	1·3768437
·344	2·7144062	·394	2·2779280	·444	1·9230678	·494	1·6261384	·544	1·3721128
0·345	2·7046573	0·395	2·2701297	0·445	1·9166226	0·495	1·6206774	0·545	1·3673948
·346	2·6949555	·396	2·2623632	·446	1·9102001	·496	1·6152333	·546	1·3626899
·347	2·6853006	·397	2·2546284	·447	1·9038002	·497	1·6098061	·547	1·3579979
·348	2·6756921	·398	2·2469251	·448	1·8974228	·498	1·6043955	·548	1·3533187
·349	2·6661296	·399	2·2392530	·449	1·8910677	·499	1·5990016	·549	1·3486523
0·350	2·6566127	0·400	2·2316119	0·450	1·8847348	0·500	1·5936243	0·550	1·3439987

k	a	k	a	k	a	k	a	k	a
0·550	1·3439987	0·600	1·1262612	0·650	0·9336939	0·700	0·7614337	0·750	0·6058600
·551	1·3393577	·601	1·1221820	·651	·9300635	·701	·7581691	·751	·6028986
·552	1·3347293	·602	1·1181127	·652	·9264412	·702	·7549111	·752	·5999427
·553	1·3301135	·603	1·1140533	·653	·9228268	·703	·7516598	·753	·5969924
·554	1·3255101	·604	1·1100038	·654	·9192204	·704	·7484150	·754	·5940475
0·555	1·3209191	0·605	1·1059641	0·655	0·9156220	0·705	0·7451767	0·755	0·5911081
·556	1·3163405	·606	1·1019342	·656	·9120315	·706	·7419450	·756	·5881742
·557	1·3117742	·607	1·0979141	·657	·9084489	·707	·7387198	·757	·5852457
·558	1·3072201	·608	1·0939037	·658	·9048742	·708	·7355010	·758	·5823227
·559	1·3026782	·609	1·0899029	·659	·9013073	·709	·7322887	·759	·5794050
0·560	1·2981485	0·610	1·0859117	0·660	0·8977481	0·710	0·7290829	0·760	0·5764927
·561	1·2936307	·611	1·0819301	·661	·8941968	·711	·7258834	·761	·5735858
·562	1·2891250	·612	1·0779581	·662	·8906531	·712	·7226903	·762	·5706842
·563	1·2846312	·613	1·0739955	·663	·8871172	·713	·7195036	·763	·5677880
·564	1·2801493	·614	1·0700424	·664	·8835889	·714	·7163232	·764	·5648970
0·565	1·2756792	0·615	1·0660987	0·665	0·8800683	0·715	0·7131491	0·765	0·5620114
·566	1·2712209	·616	1·0621643	·666	·8765553	·716	·7099813	·766	·5591310
·567	1·2667743	·617	1·0582393	·667	·8730498	·717	·7068198	·767	·5562558
·568	1·2623394	·618	1·0543236	·668	·8695519	·718	·7036645	·768	·5533859
·569	1·2579161	·619	1·0504171	·669	·8660616	·719	·7005154	·769	·5505213
0·570	1·2535043	0·620	1·0465198	0·670	0·8625787	0·720	0·6973725	0·770	0·5476618
·571	1·2491040	·621	1·0426317	·671	·8591032	·721	·6942358	·771	·5448075
·572	1·2447151	·622	1·0387527	·672	·8556352	·722	·6911053	·772	·5419584
·573	1·2403377	·623	1·0348828	·673	·8521746	·723	·6879808	·773	·5391144
·574	1·2359715	·624	1·0310219	·674	·8487214	·724	·6848625	·774	·5362756
0·575	1·2316167	0·625	1·0271701	0·675	0·8452755	0·725	0·6817503	0·775	0·5334419
·576	1·2272730	·626	1·0233272	·676	·8418370	·726	·6786441	·776	·5306132
·577	1·2229405	·627	1·0194933	·677	·8384057	·727	·6755440	·777	·5277897
·578	1·2186192	·628	1·0156683	·678	·8349816	·728	·6724499	·778	·5249712
·579	1·2143089	·629	1·0118521	·679	·8315648	·729	·6693618	·779	·5221578
0·580	1·2100097	0·630	1·0080447	0·680	0·8281552	0·730	0·6662796	0·780	0·5193494
·581	1·2057214	·631	1·0042461	·681	·8247528	·731	·6632035	·781	·5165460
·582	1·2014440	·632	1·0004563	·682	·8213575	·732	·6601332	·782	·5137476
·583	1·1971774	·633	0·9966752	·683	·8179694	·733	·6570689	·783	·5109542
·584	1·1929217	·634	0·9929027	·684	·8145883	·734	·6540104	·784	·5081658
0·585	1·1886768	0·635	0·9891389	0·685	0·8112143	0·735	0·6509579	0·785	0·5053823
·586	1·1844426	·636	·9853837	·686	·8078473	·736	·6479112	·786	·5026037
·587	1·1802190	·637	·9816371	·687	·8044874	·737	·6448703	·787	·4998301
·588	1·1760060	·638	·9778989	·688	·8011344	·738	·6418352	·788	·4970613
·589	1·1718036	·639	·9741693	·689	·7977884	·739	·6388059	·789	·4942975
0·590	1·1676118	0·640	0·9704481	0·690	0·7944493	0·740	0·6357824	0·790	0·4915385
·591	1·1634304	·641	·9667354	·691	·7911171	·741	·6327646	·791	·4887843
·592	1·1592594	·642	·9630310	·692	·7877918	·742	·6297526	·792	·4860350
·593	1·1550988	·643	·9593350	·693	·7844733	·743	·6267462	·793	·4832905
·594	1·1509485	·644	·9556473	·694	·7811617	·744	·6237456	·794	·4805508
0·595	1·1468085	0·645	0·9519679	0·695	0·7778568	0·745	0·6207506	0·795	0·4778159
·596	1·1426788	·646	·9482967	·696	·7745587	·746	·6177612	·796	·4750858
·597	1·1385592	·647	·9446338	·697	·7712674	·747	·6147775	·797	·4723604
·598	1·1344498	·648	·9409790	·698	·7679828	·748	·6117994	·798	·4696398
·599	1·1303505	·649	·9373324	·699	·7647049	·749	·6088269	·799	·4669239
0·600	1·1262612	0·650	·9336939	0·700	0·7614337	0·750	0·6058600	0·800	0·4642128

k	a	k	a	k	a	k	a
0·800	0·4642128	0·850	0·3343447	0·900	0·2145557	0·950	0·1034788
·801	·4615063	·851	·3318552	·901	·2122529	·951	·1013381
·802	·4588045	·852	·3293697	·902	·2099535	·952	·0992005
·803	·4561073	·853	·3268882	·903	·2076576	·953	·0970658
·804	·4534148	·854	·3244107	·904	·2053652	·954	·0949342
0·805	0·4507270	0·855	0·3219371	0·905	0·2030762	0·955	0·0928056
·806	·4480438	·856	·3194675	·906	·2007906	·956	·0906799
·807	·4453652	·857	·3170019	·907	·1985084	·957	·0885573
·808	·4426911	·858	·3145402	·908	·1962296	·958	·0864376
·809	·4400217	·859	·3120824	·909	·1939542	·959	·0843209
0·810	0·4373568	0·860	0·3096286	0·910	0·1916824	0·960	0·0822071
·811	·4346965	·861	·3071786	·911	·1894137	·961	·0800963
·812	·4320407	·862	·3047325	·912	·1871484	·962	·0779885
·813	·4293894	·863	·3022903	·913	·1848866	·963	·0758836
·814	·4267426	·864	·2998520	·914	·1826281	·964	·0737816
0·815	0·4241004	0·865	0·2974176	0·915	0·1803729	0·965	0·0716825
·816	·4214626	·866	·2949869	·916	·1781211	·966	·0695864
·817	·4188292	·867	·2925601	·917	·1758725	·967	·0674932
·818	·4162003	·868	·2901372	·918	·1736274	·968	·0654028
·819	·4135759	·869	·2877180	·919	·1713855	·969	·0633154
0·820	0·4109558	0·870	0·2853027	0·920	0·1691469	0·970	0·0612308
·821	·4083402	·871	·2828911	·921	·1669116	·971	·0591492
·822	·4057290	·872	·2804833	·922	·1646795	·972	·0570704
·823	·4031222	·873	·2780793	·923	·1624508	·973	·0549944
·824	·4005197	·874	·2756790	·924	·1602253	·974	·0529213
0·825	0·3979215	0·875	0·2732825	0·925	0·1580031	0·975	0·0508511
·826	·3953278	·876	·2708897	·926	·1557841	·976	·0487837
·827	·3927383	·877	·2685007	·927	·1535683	·977	·0467191
·828	·3901532	·878	·2661153	·928	·1513558	·978	·0446574
·829	·3875723	·879	·2637337	·929	·1491465	·979	·0425985
0·830	0·3849957	0·880	0·2613557	0·930	0·1469404	0·980	0·0405424
·831	·3824235	·881	·2589815	·931	·1447374	·981	·0384891
·832	·3798554	·882	·2566109	·932	·1425377	·982	·0364386
·833	·3772916	·883	·2542439	·933	·1403412	·983	·0343909
·834	·3747321	·884	·2518806	·934	·1381478	·984	·0323460
0·835	0·3721768	0·885	0·2495210	0·935	0·1359576	0·985	0·0303038
·836	·3696257	·886	·2471649	·936	·1337706	·986	·0282644
·837	·3670787	·887	·2448125	·937	·1315867	·987	·0262278
·838	·3645360	·888	·2424637	·938	·1294060	·988	·0241939
·839	·3619974	·889	·2401185	·939	·1272283	·989	·0221628
0·840	0·3594630	0·890	0·2377769	0·940	0·1250538	0·990	0·0201345
·841	·3569327	·891	·2354389	·941	·1228825	·991	·0181088
·842	·3544066	·892	·2331044	·942	·1207142	·992	·0160859
·843	·3518846	·893	·2307735	·943	·1185490	·993	·0140657
·844	·3493667	·894	·2284461	·944	·1163869	·994	·0120482
0·845	0·3468528	0·895	0·2261222	0·945	0·1142279	0·995	0·0100335
·846	·3443431	·896	·2238019	·946	·1120720	·996	·0080214
·847	·3418374	·897	·2214851	·947	·1099191	·997	·0060120
·848	·3393358	·898	·2191718	·948	·1077693	·998	·0040053
·849	·3368383	·899	·2168620	·949	·1056226	·999	·0020013
0·850	0·3343447	0·900	0·2145557	0·950	0·1034788	1·000	0·0000000

Biometrika (1960), **47**, 3 *and* 4, *p.* 447
Printed in Great Britain

Miscellanea

Studies in the history of probability and statistics. X. Where shall the history of statistics begin?

By M. G. KENDALL
Research Techniques Division, London School of Economics and Political Science

1. A history must start somewhere, but history has no beginning. It is therefore natural, or at least fashionable, for the historian of any particular period or of any particular subject to preface his main treatment by tracing the roots of his theme back into the past as far as he possibly can. The writer on any modern idea who can claim that the Chinese thought of it first in the Shang period is usually regarded as having scored a point. The underlying assumption that historical development is continuous is sometimes justifiable and sometimes, though not justifiable, innocuous. A history of marine insurance from the foundation of Edward Lloyd's coffee house in 1712 may reasonably begin with an account of the sea-laws of Rhodes of two thousand years earlier. There is no great harm in a history of Newmarket race course glancing at Eohippus and passing lightly over the chariot races of the ancient world before lining up at the starting post of A.D. 1640.

2. But there are dangers in pursuing the roots of a subject down to its slenderest fibrils. A word, an outward form, a material expression may be continuous whereas the concepts which it embodies may drastically change. In earlier articles in this series Dr David and I have tried to emphasize that, although the instruments of gaming had existed for several thousand years, probability theory as a conceptual abstraction of the laws of chance did not come into being until the sixteenth century. In this present note I want to stress the fact that something of the same kind is true of what we call 'statistics'.

3. The temptation to begin a history of statistics with references to endeavours in the ancient world to record information about states is one which no writer has been able to resist. The numbering of the people of Israel, Augustus's balance sheets of the Roman Empire, Charlemagne's inventory of his possessions, the Doomsday book, are felt to be a natural preliminary to the emergence of descriptive statistics in the eighteenth century; as though they were early attempts undertaken in the spirit of modern statistics, hampered by the backward state of social organization, but inspired by the same concepts.

4. To believe this, in my opinion, is to fail to understand either the basis of the statistical approach or the nature of the statistical method. In this note I put forward the thesis that statistics as we now understand the term did not commence until the seventeenth century, and then not in the field of 'statistics'.

5. The early inventories, of which the Doomsday book is an outstanding example, were not, of course, called statistical. The first use of the word known to me occurs in a work by an Italian historian Girolamo Ghilini, who in 1589 refers to an account of the *civile, politica, statistica e militare scienza*. From then onwards we have a series of well-known landmarks in the history of political description: Sansovino's (1561) *Dal governo ed amministrazione dei diversi regie e republiche*; Botero's (1650) *Le relazione universale*; Seckendorf's (1656) *Teutscher Fürstenstaat*; Oldenburger's (1673) publication of Conring's lectures; Anchersen's (1741) *Statuum Cultiorum in Tabulis*; Süssmilch's (1741) *Göttliche Ordnung*; Achenwall's (1748) Public Dissertation; Büsching's (1754) *Neue Erdbeschreibung*. These works are often referred to as 'statistical' and their authors used the word to denote the method of their studies. But they are entirely concerned with the description of political states and any numerical information in them appears by accident or convenience. They are the forerunners of our geographical and political works of reference like *The Statesman's Yearbook*, containing an immense amount of interesting information, but 'statistical' only in the sense of relating to what the Germans called *Staatenkunde*. The point has been well made by Westergaard (1932) in his modest and scholarly history. If it were not, he says, for the fact that the same word 'statistics' is used to denote both this early work and statistics in the modern sense, the student of the latter would probably pay no attention to the former; and Westergaard implies that he would lose little by ignoring it, a judgment with which I agree. The true ancestor of modern statistics is not seventeenth-century statistics but Political Arithmetic.

6. There is another reason, of quite a different kind, for not seeking the origins of modern statistics too curiously among the ruins of the medieval world. The early inventories of the Doomsday type were

pictures of a stationary state of society. 'The feudal state of the Middle Ages', says Burckhardt (1929), 'knew of nothing more than categories of signorial rights and possessions. It looked upon production as a fixed quantity, which it approximately is, so long as we have to do with landed property only'. In those days the world was a small affair, only a few thousand years old, the fixed centre of the universe, guided under a rigid set of rules by state and church. There was no particular point in trying to keep track of minor changes in such a system. The main motives for statistical inquiry, the estimation of military man-power and of taxation revenue, were absent. Not that soldiers and taxes were unimportant, but that the state's resources could be estimated simply on the basis of what had resulted from the last levy.

7. For the dissolution of this system we look naturally to those states which first began to found their economy in trade and manufacture; and we find, accordingly, that in fourteenth- and fifteenth-century Italy statistics in the sense of political arithmetic began to emerge. There are statistical accounts of some of the Italian city-states as early as the thirteenth century—there have survived accounts of Asti in A.D. 1250 and of Milan in A.D. 1288. But these are inventories, although greater emphasis on goods, shops and persons is already noticeable. To Venice, apparently, goes the honour of priority in a new approach. It was the Venetians who first counted the population as souls, that is to say, included all living individuals and did not, for example, confine themselves to citizens who were capable of bearing arms or paying taxes. It was at Venice on one occasion, too, that a proposal to go to war with a neighbouring state was rejected on the economic grounds that fighting the customer was bad for both sides. This general interest in numerical assessment was echoed at Florence and, a little later, among some of the despotisms of northern Italy.

8. Bearing in mind then, that fifteenth-century Italy saw striking developments in accountancy and mathematics, and hence that the necessary technical equipment was available we may conclude that the commercial requirements of the age alone necessitated the collection of a great deal of material which would nowadays be called descriptive statistics. But we are, I think, still short of a statistical approach. Counting was by complete enumeration and still tended to be a record of a situation rather than a basis for estimation or prediction in an expanding economy. In any case the Risorgimento occurred too soon for a Europe which was to spend many years fighting destructive wars at home and colonizing an unknown continent abroad. There appears to have been very little accomplished outside Italy either in descriptive statistics or in probability in the sixteenth century. Not until Europe had settled down about the time of our own restoration (A.D. 1660) did political arithmetic (including life assurance) begin in earnest.

9. Progress was then rapid, and as in many scientific developments, the basic ideas seem to arise independently at many points in Western Europe. John Graunt's famous *Observations* appeared in 1662. Hudde's book on *Annuities* appeared in 1671. Ludwig Huyghens writes to his brother Christian in 1669: 'I have just been making a table showing how long people of given age have to live. . . .Live well! According to my calculations you will live to be about 56½ and I to 55.' At the close of the century we have, within a few years, Petty's *Political Arithmetic* (1690), Halley's *Estimate* (1693), Gregory King's *Observations* (1696), Davenant's *Discourses on the Public Revenues* (1698) and, in 1703, the first census of modern times, in Iceland. There are many accounts of this work, especially of Graunt's and Petty's. Anyone who reads Graunt or King must, in my view, feel that this is the point at which statistics really began. The whole attack on the problems can be appreciated by a statistician or a demographer nowadays. These men, one feels, thought as we think today, with most imperfect material but with enviable penetration. They *reasoned* about their data.

10. This is not to imply that statistics in the old sense came suddenly to an end, or that the inferential type of approach spread from demographic applications to other fields with any speed. The political description of states continued and flourished, becoming more numerical as additional data were collected, but remaining essentially the systematic collection of facts. During the eighteenth century we find the two subjects running side by side with very little apparent connexion between them. In the one, enumerative surveys such as Sinclair's (1793) *Statistical Account of Scotland*, and discourses by men such as Wargentin, Büsching and Schlöger, whose very names are unknown to modern statistics; in the other, contributions to demography and actuarial science by men such as Demoivre, Daniel Bernoulli, Laplace, Euler and Poisson. It was not until the middle of the nineteenth century that these two branches of knowledge showed any sign of coalescence.

11. I conclude then, that statistics in any sense akin to our own cannot be traced back before about A.D. 1660. To look backwards from that point is interesting but can be misleading, unless its purpose is, not to show that modern statistical ideas can be traced back into medieval and ancient times, but that such ideas were almost unknown prior to A.D. 1660. There *are* new things under the sun and the attitude of mind towards experimental knowledge about natural and social phenomena which flourished after the Restoration is one of them.

REFERENCES

BURCKHARDT, J. (1929). *The Civilization of the Renaissance in Italy* (translated from the 15th edition by S. G. C. Middlemore. Harrap.

WESTERGAARD, H. (1932). *Contributions to the History of Statistics*. London: P. S. King & Son.

On the number of renewals in a random interval

BY D. R. COX

Birkbeck College, London

1. INTRODUCTION

There have been many studies of the number N_t of renewals occurring in a fixed time interval $(0, t)$ of a renewal process; see especially Feller (1957, Chapter 13) and Smith (1958, 1959). Suppose, however, that we are interested in the number N of renewals in an interval $(0, T)$, where T is a random variable distributed independently of the renewal process. For example, T may be the lifetime of a whole piece of equipment, and we may be concerned with the number of component replacements that are necessary during T.

The object of this note is to point out that when the distribution of T is of the gamma type, the properties of N can be found without the inversion of Laplace transforms. Processes in continuous time only are considered.

2. FORMULATION OF PROBLEM

A renewal process is defined by a sequence $(X_1, X_2, \ldots)$ of independent identically distributed random variables, taking only positive values, and assumed here to have a continuous distribution with probability density function (p.d.f.) $f(x)$. The random variable X_i can be interpreted as the lifetime of the ith component, so that renewals occur at times

$$X_1, X_1+X_2, \ldots. \tag{1}$$

The random variable N_t is defined as the number of renewals in $(0, t)$, i.e. formally as the number of partial sums (1) in the interval $(0, t)$. One of the especial interests of renewal processes is that they arise not only in connexion with the renewal of components, but also in a wide range of problems of applied probability theory (Smith, 1958).

If
$$p_r(t) = \text{prob}\,(N_t = r) \tag{2}$$

and
$$G(z, t) = \sum_{r=0}^{\infty} z^r p_r(t), \tag{3}$$

then it is known that the Laplace transforms of these functions are

$$p_r^*(s) = [f^*(s)]^r \frac{1-f^*(s)}{s}, \tag{4}$$

$$G^*(z, s) = \frac{1-f^*(s)}{s(1-zf^*(s))}. \tag{5}$$

Here the asterisk denotes a Laplace transform with respect to t, so that, for example,

$$f^*(s) = \int_0^{\infty} e^{-st} f(t)\, dt.$$

From (5) the Laplace transform of the mean, and more generally of the factorial moments of N_t, can be found by differentiating with respect to z at $z = 1$. Alternatively, and more neatly, the ϕ-moments of Smith (1959) are defined by

$$\phi_r(t) = E[(N_t+1)(N_t+2)\ldots(N_t+r)], \tag{6}$$

and for these we have the simple formula that

$$\phi_r^*(s) = \frac{r!}{[1-f^*(s)]^r}. \tag{7}$$

In general the explicit inversion of these results into usable formulae for $p_r(t)$ and $G(z,t)$ is possible only for special $f(x)$. Simple general results are, however, available for large t. The asymptotic normality of N_t was proved by Feller (1957, Chapter 13) and Smith (1959) has shown under weak assumptions that the rth cumulant of N_t has, for large t, the form

$$a_r t + b_r + o(1) \tag{8}$$

and has given a_r, b_r ($r \leqslant 8$) in terms of the moments of the X's.

Now suppose that the positive random variable T is distributed independently of $(X_1, X_2, \ldots)$ and has p.d.f. $q(t)$. Let N be the number of renewals in $(0, T)$. Then the probability generating function of N is given by

$$G(z) \equiv \sum_{r=0}^{\infty} p_r z^r = \int_0^{\infty} G(z,t)\, q(t)\, dt. \tag{9}$$

We consider here how this can be obtained without inverting the Laplace transform (5).

3. A gamma distribution for T

We now suppose that the p.d.f. $q(t)$ is of the Γ type, or more generally is a weighted sum of terms of the Γ type. Take for $q(t)$ the special form

$$q_{k,m}(t) = \frac{(k/m)^k\, t^{k-1}\, e^{-kt/m}}{(k-1)!}, \tag{10}$$

where m is the mean of T and the coefficient of variation of T is $1/\sqrt{k}$.

Now

$$G^*(z,s) = \int_0^{\infty} G(z,t)\, e^{-st} dt,$$

so that

$$\left(-\frac{\partial}{\partial s}\right)^{k-1} G^*(z,s) = \int_0^{\infty} G(z,t)\, t^{k-1} e^{-st} dt. \tag{11}$$

This is immediate for integral k; to justify the formula for general k we can interpret $(-\partial/\partial s)^{k-1}$ by using fractional derivatives in the sense of Liouville; see, for example, Courant (1946, p. 339).

It follows from (11) that

$$\begin{aligned} G(z) &= \int_0^{\infty} G(z,t)\, q_{k,m}(t)\, dt \\ &= \frac{(k/m)^k}{(k-1)!} \left(-\frac{\partial}{\partial s}\right)^{k-1} [G^*(z,s)]_{s=k/m} \qquad (12) \\ &= \frac{(k/m)^k}{(k-1)!} \left(-\frac{\partial}{\partial s}\right)^{k-1} \left[\frac{1-f^*(s)}{s(1-zf^*(s))}\right]_{s=k/m}. \qquad (13) \end{aligned}$$

By extracting the coefficient of z^r, or by arguing directly from (4), we have that p_r, the probability that N equals r, is

$$p_r = \frac{(k/m)^k}{(k-1)!} \left(-\frac{\partial}{\partial s}\right)^{k-1} \left[\frac{1-f^*(s)}{s} (f^*(s))^r\right]_{s=k/m}. \tag{14}$$

The general result is that if $w(t)$ is a function of t, then $w_{k,m}$, the average of $w(t)$ with respect to the p.d.f. $q_{k,m}(t)$ is given by

$$w_{k,m} = \frac{(k/m)^k}{(k-1)!} \left(-\frac{\partial}{\partial s}\right)^{k-1} [w^*(s)]_{s=k/m}. \tag{15}$$

An interesting special case is obtained by letting $k \to \infty$. Then the gamma distribution (10) tends to the point concentration of probability at m, and therefore $w_{k,m}$ tends to $w(m)$. If we replace m by t, we have that

$$w(t) = \lim_{k\to\infty} \frac{(k/t)^k}{(k-1)!} \left(-\frac{d}{ds}\right)^{k-1} [w^*(s)]_{s=k/t}, \tag{16}$$

an inversion formula for the Laplace transform due to Post (1930).

4. Some special cases

The evaluation of (13) and (14) is easiest when k is a small integer, that is when T is distributed proportionally to chi-squared with a small even number of degrees of freedom, a so-called Erlangian

distribution. In particular if T is exponentially distributed, $k = 1$, and then

$$G(z) = \frac{1-f^*(1/m)}{1-zf^*(1/m)}, \tag{17}$$

so that the distribution of N is geometric with mean $f^*(1/m)/(1-f^*(1/m))$. If m is large and if only the low-order moments of the X's are known, $f^*(1/m)$ can be approximated by expansion in series. The intuitive explanation of the occurrence of a geometric distribution is as follows. Consider a Poisson process of rate $1/m$ producing events called 'catastrophes'. The probability that a lifetime X_i contains no catastrophe is easily shown to be $f^*(1/m)$, the probabilities for different X_i being independent. The random variable N can thus be interpreted as the number of lifetimes that are completed before a catastrophe, and has the form of the number of 'failures' recorded before the first 'success'. The application of a similar idea to interpret other equations in renewal theory has been outlined by Cox (1957).

The result (17) generalizes simply if the distribution of T is a linear combination of exponential terms; such a combination is sometimes useful, for example, to represent a distribution with coefficient of variation greater than one.

If the distribution of the X's also is of the gamma type, with p.d.f.

$$\frac{(\kappa/\mu)^\kappa x^{\kappa-1} e^{-\kappa x/\mu}}{(\kappa-1)!}, \tag{18}$$

then

$$f^*(s) = (1+\mu s/\kappa)^{-\kappa}$$

and

$$G(z) = \frac{(k/m)^k}{(k-1)!}\left(-\frac{\partial}{\partial s}\right)^{k-1}\left[\frac{(1+\mu s/\kappa)^\kappa - 1}{s(1+\mu s/\kappa)^\kappa - sz}\right]_{s=k/m}. \tag{19}$$

In this formula the symbols k, m refer to the distribution of T, whereas κ, μ are the corresponding parameters of the distribution of the X's.

If $\kappa = 1$ the renewals form a Poisson process and (19) then gives

$$G(z) = \frac{(k\mu/m)^k}{(1+k\mu/m-z)^k}, \tag{20}$$

a negative binomial distribution. This result is equivalent to the standard result that a Poisson distribution compounded with a gamma distribution leads to a negative binomial distribution. For, given the value of $T = t$, say, N_t has a Poisson distribution with mean μt, and (20) gives the distribution of N when T, and hence μT, follows a gamma distribution.

If $k = \kappa = 2$

$$G(z) = \frac{(\mu/m)^2(\mu/m+3)(\mu/m+1)}{[(1+\mu/m)^2-z]^2} + \frac{(\mu/m)^2 z}{[(1+\mu/m)^2-z]^2}, \tag{21}$$

so that the distribution of N is a weighted sum of a negative binomial distribution of index two and of the same distribution with all the variate-values increased by one. In a particular numerical application, (19) can be used to obtain the distribution of N provided k and κ are not large; if the small values of N are of particular interest, it may be easier to work from (14).

For the ϕ-moments of N, we have from (7) and (15) that

$$\begin{aligned}\phi_r &= E[(N+1)(N+2)\ldots(N+r)] \\ &= \frac{(k/m)^k}{(k-1)!}\left(-\frac{d}{ds}\right)^{k-1}\left[\frac{r!}{(1-f^*(s))^r}\right]_{s=k/m},\end{aligned}$$

and in particular, if $f(x)$ is given by (18),

$$\begin{aligned}\phi_1 &= E(N+1) \\ &= \frac{(k/m)^k}{(k-1)!}\left(-\frac{d}{ds}\right)^{k-1}\left[\frac{(1+\mu s/\kappa)^\kappa}{(1+\mu s/\kappa)^\kappa - 1}\right]_{s=k/m},\end{aligned} \tag{22}$$

from which explicit expressions can again be obtained, especially if k is a small integer. If m is large and k small, all the values of T will be large, and the following special cases of (8) can be used

$$E(N_t|t) = \frac{t}{\mu} + \left(\frac{\mu_2}{2\mu^2} - 1\right) + o(1),$$

$$V(N_t|t) = \frac{\mu_2-\mu_1^2}{\mu^3}t + \left(\frac{5\mu_2^2}{4\mu^4} - \frac{2\mu_3}{3\mu^3} - \frac{\mu_2}{2\mu_2}\right) + o(1),$$

where μ_r is the rth moment of the X's about the origin and $\mu_1 = \mu$.

Then
$$E(N) = \frac{m}{\mu} + \left(\frac{\mu^2}{2\mu^2} - 1\right) + o(1), \tag{23}$$

$$V(N) = \frac{(\mu_2 - \mu^2)\, m}{\mu^3} + \frac{m^2}{k\mu^2} + \left(\frac{5\mu_2^2}{4\mu^4} - \frac{2\mu_3}{3\mu^3} - \frac{\mu^2}{2\mu^2}\right) + o(1). \tag{24}$$

5. Equilibrium renewal process

The results above apply to a renewal process in which the time-origin is the start of the life of a component, i.e. in which the random variable X_1 has the same distribution as the other X's. In applications it often happens that X_1 has a different distribution from the other X's, and an important special case is the so-called equilibrium renewal process (Cox & Smith, 1954) in which X_1 has the p.d.f.

$$\left(1 - \int_0^x f(u)\, du\right) \Big/ \mu; \tag{25}$$

this corresponds to a system in which the process has been running for a long time before the time-origin. If N_t is the number of renewals in $(0, t)$ and $G_e(z, t)$ its probability generating function, then analogously to (5)

$$G_e^*(z, s) = \frac{1}{s} - \frac{(1 - f^*(s))\,(1 - z)}{s^2 \mu (1 - z f^*(s))}, \tag{26}$$

and the previous arguments apply. In particular if T has an exponential distribution, the result analogous to (17) is that

$$G_e(z) = 1 - \frac{(1 - z)\,(1 - f^*(1/m))}{\frac{\mu}{m}(1 - z f^*(1/m))}. \tag{27}$$

References

Courant, R. (1936). *Differential and Integral Calculus*, 2. London: Blackie.

Cox, D. R. (1957). Contribution to discussion of paper by J. G. Skellam and L. R. Shenton. *J. R. Statist. Soc. B*, 113–14.

Cox, D. R. & Smith, W. L. (1954). On the superposition of renewal processes. *Biometrika*, **41**, 91–9.

Feller, W. (1957). *Introduction to Probability Theory and its Applications*, 2nd ed. New York: Wiley.

Post, E. L. (1930). Generalised differentiation. *Trans. Amer. Math. Soc.* **32**, 723–81.

Smith, W. L. (1958). Renewal theory and its ramifications. *J. R. Statist. Soc.* B, **20**, 243–302.

Smith, W. L. (1959). On the cumulants of renewal processes. *Biometrika*, **46**, 1–29.

A note on sufficiency in regular Markov chains

By B. R. Bhat and J. Gani†

University of California, Berkeley and University of Western Australia, Nedlands

1. Introduction and summary

For positively regular (p.r.) Markov chains with a finite number s of states and transition probabilities $p_{ij}(\theta)$ $(i, j = 1, 2, \ldots, s)$ defined by an unknown parameter θ, Gani (1955, 1956) has shown that in realizations starting from a fixed state E_α $(\alpha = 1, 2, \ldots, s)$ the form of the $p_{ij}(\theta)$ admitting a sufficient estimator of θ is

$$p_{ij}(\theta) = A_{ij} \exp\{K_{ij} \Lambda_1(\theta) + \lambda_2(\theta)\}, \tag{1·1}$$

where the A_{ij}, K_{ij} are constants. Since $\sum_{j=1}^{s} p_{ij}(\theta) = 1$ for all rows $i = 1, 2, \ldots, s$, it follows that for every i

$$\exp\{-\lambda_2(\theta)\} = \sum_{j=1}^{s} A_{ij} \exp\{K_{ij} \Lambda_1(\theta)\}; \tag{1·2}$$

this means that there can exist only $r \leqslant s$ distinct exponents K_{ij} such that apart from coefficients A_{ij},

† This research was supported in part by the Office of Naval Research, at Columbia University, New York.

the $p_{ij}(\theta)$ in each row of the stochastic matrix consist of the same $r \leqslant s$ distinct forms of type (1·1) in varying orders.

In this note, the initial state E_α is random, and a slightly more general result is obtained for p.r. chains with initial probabilities $a_\alpha(\theta)$ and transition probabilities $p_{ij}(\theta)$ $(\alpha, i, j = 1, 2, \ldots, s)$. A special regular (r.) but not p.r. chain is also considered, and some properties and distributions of the sufficient estimators of θ in both p.r. and r. chains are given.

2. The p.r. Markov chain

Let S be a realization of n transitions, starting with a random state E_α, from a p.r. Markov chain whose initial and transition probabilities are $a_\alpha(\theta)$, $p_{ij}(\theta)$ $(\alpha, i, j = 1, 2, \ldots, s)$, respectively. For such a realization the log-likelihood function is given by

$$L(S) = \sum_{\alpha=1}^{s} x_\alpha \ln a_\alpha(\theta) + \sum_{i,j=1}^{s} n_{ij} \ln p_{ij}(\theta), \tag{2·1}$$

where each x_α may be 0 or 1 but

$$\sum_{\alpha=1}^{s} x_\alpha = x = 1,$$

and the n_{ij} are transition frequencies such that

$$\sum_{i,j=1}^{s} n_{ij} = n.$$

Let $T(x, n)$ be a sufficient estimator of θ, where $\mathbf{x}, \mathbf{n}$ indicate the elements $\{x_\alpha\}$ and $\{n_{ij}\}$; the factorizability condition then leads to the equation

$$L(S) = f(\mathbf{x}, \mathbf{n}) + F(T, \theta; x, n), \tag{2·2}$$

both x and n being constants which may (but need not necessarily) appear explicitly in $L(S)$.

Assuming differentiability with respect to θ for all $a_\alpha(\theta)$, $p_{ij}(\theta)$, it follows from (2·1) and (2·2) that

$$\partial F/\partial\theta = G(T, \theta; x, n) = \sum_{\alpha=1}^{s} x_\alpha d\{\ln a_\alpha(\theta)\}/d\theta + \sum_{i,j=1}^{s} n_{ij} d\{\ln p_{ij}(\theta)\}/d\theta, \tag{2·3}$$

where G is a linear function of the $\{x_\alpha\}$ and $\{n_{ij}\}$. For x and n to appear explicitly in (2·3), we must have

$$\left.\begin{aligned} d\{\ln a_\alpha(\theta)\}/d\theta &= u_\alpha(\theta) + u(\theta), \\ d\{\ln p_{ij}(\theta)\}/d\theta &= w_{ij}(\theta) + w(\theta), \end{aligned}\right\} \quad (\alpha, i, j = 1, 2, \ldots, s) \tag{2·4}$$

the function G then being of the form

$$G(T, \theta; x, n) = \sum_{\alpha=1}^{s} x_\alpha u_\alpha(\theta) + x u(\theta) + \sum_{i,j=1}^{s} n_{ij} w_{ij}(\theta) + n w(\theta). \tag{2·5}$$

Although the $\{x_\alpha\}$, $\{n_{ij}\}$ are discrete, the function G is formally differentiable with respect to each x_α, n_{ij} as well as x, n; if we further assume that T is differentiable with respect to the $\{x_\alpha\}$, $\{n_{ij}\}$, we readily obtain

$$\left.\begin{aligned} u_\alpha(\theta) &= \{\partial G/\partial T\}\{\partial T/\partial x_\alpha\}, \\ w_{ij}(\theta) &= \{\partial G/\partial T\}\{\partial T/\partial n_{ij}\}, \end{aligned}\right\} \quad (\alpha, i, j = 1, 2, \ldots, s). \tag{2·6}$$

From this, it follows that $\partial T/\partial x_\alpha$ and $\partial T/\partial n_{ij}$ must necessarily be functions $U_\alpha^{-1}(T)$, $W_{ij}^{-1}(T)$ of T, respectively, not depending on x_α and n_{ij}. This means that

$$\partial G/\partial T = u_\alpha(\theta)\, U_\alpha(T) = w_{ij}(\theta)\, W_{ij}(T) \quad (\alpha, i, j = 1, 2, \ldots, s), \tag{2·7}$$

and this will hold only if for all $\alpha, i, j = 1, 2, \ldots, s$,

$$\left.\begin{aligned} u_\alpha(\theta)\, K_\alpha^{-1} &= w_{ij}(\theta)\, K_{ij}^{-1} = k(\theta), \\ U_\alpha(T)\, K_\alpha &= W_{ij}(T)\, K_{ij} = v(T), \end{aligned}\right\} \tag{2·8}$$

where the K_α, K_{ij} are constants.

Integration of (2·4) leads directly to the exponential forms for the initial and transition probabilities

$$\left.\begin{aligned} a_\alpha(\theta) &= A_\alpha \exp\{K_\alpha \Lambda_1(\theta) + \Lambda_2(\theta)\}, \\ p_{ij}(\theta) &= A_{ij} \exp\{K_{ij} \Lambda_1(\theta) + \lambda_2(\theta)\}, \end{aligned}\right\} \quad (\alpha, i, j = 1, 2, \ldots, s), \tag{2·9}$$

where A_α, A_{ij} are constants, and $\Lambda_1(\theta) = \int k(\theta)\,d\theta$; it should be noted that the functions $\Lambda_2(\theta) = \int u(\theta)\,d\theta$ and $\lambda_2(\theta) = \int w(\theta)\,d(\theta)$ are distinct. Since the form of the $p_{ij}(\theta)$ is identical with (1·1), this clearly leads to a similar matrix configuration for the transition probabilities. It is obvious that since $\sum_{\alpha=1}^{s} a_\alpha(\theta) = 1$, then

$$\exp\{-\Lambda_2(\theta)\} = \sum_{\alpha=1}^{s} A_\alpha \exp\{K_\alpha \Lambda_1(\theta)\}. \tag{2·10}$$

The function G is given by

$$G(T, \theta; x, n) = k(\theta)\, g(T) + xu(\theta) + nw(\theta), \tag{2·11}$$

where $g(T) = \int v(T)\,dT$ is

$$g(T) = \int v(T) \left\{ \sum_{\alpha=1}^{s} (\partial T/\partial x_\alpha)\,dx_\alpha + \sum_{i,j=1}^{s} (\partial T/\partial n_{ij})\,dn_{ij} \right\} = \sum_{\alpha=1}^{s} x_\alpha K_\alpha + \sum_{i,j=1}^{s} n_{ij} K_{ij}, \tag{2·12}$$

a linear function of the x_α and n_{ij}.

3. A special r. Markov chain

Consider now the r. Markov chain with $s+1$ states whose initial and transition probabilities are $a_\alpha(\theta)$, $p_{ij}(\theta)$ $(\alpha, i, j = 0, 1, \ldots, s)$, respectively, but such that $a_0(\theta) = 0$ and $p_{00}(\theta) = 1$, the state 0 being the only absorbing state in the chain. If S' is a realization in $n+1$ transitions from such a chain (n now being a random variable) starting from a random state $\alpha = 1, 2, \ldots, s$, the log-likelihood function is given by

$$L(S') = \sum_{\alpha=1}^{s} x_\alpha \ln a_\alpha(\theta) + \sum_{i,j=1}^{s} n_{ij} \ln p_{ij}(\theta) + \sum_{\beta=1}^{s} y_\beta \ln p_{\beta 0}(\theta), \tag{3·1}$$

where each x_α and y_β may be 0 or 1 but $\sum_{\alpha=1}^{s} x_\alpha = x = 1$ and $\sum_{\beta=1}^{s} y_\beta = y = 1$, the n_{ij} being transition frequencies such that $\sum_{i,j=1}^{s} n_{ij} = n$.

Let $T(x, n, y)$ be a sufficient estimator of θ, where $\mathbf{x}, \mathbf{n}, \mathbf{y}$ represent the elements $\{x_\alpha\}$, $\{n_{ij}\}$, $\{y_\beta\}$, respectively. The factorizability condition for sufficiency will then give

$$L(S') = f(\mathbf{x}, \mathbf{n}, \mathbf{y}) + F(T, \theta; x, y), \tag{3·2}$$

both x and y being constants which may appear explicitly in $L(S')$. It should be noted that in contrast with (2·2), n which is now a random variable does not appear in the function F.

It follows, exactly as in § 2, that

$$\partial F/\partial\theta = G(T, \theta; x, y) = \sum_{\alpha=1}^{s} x_\alpha u_\alpha(\theta) + xu(\theta) + \sum_{i,j=1}^{s} n_{ij} w_{ij}(\theta) + \sum_{\beta=1}^{s} y_\beta v_\beta(\theta) + yv(\theta), \tag{3·3}$$

where $d\{\ln a_\alpha(\theta)\}/d\theta = u_\alpha(\theta) + u(\theta)$ and $d\{\ln p_{\beta 0}(\theta)\}/d\theta = v_\beta(\theta) + v(\theta)$, but $d\{\ln p_{ij}(\theta)\}/d\theta = w_{ij}(\theta)$ only, for $\alpha, \beta, i, j = 1, 2, \ldots, s$. Following some analysis similar to that in § 2, it is readily found that

$$\left.\begin{aligned} a_\alpha(\theta) &= A_\alpha \exp\{K_\alpha \Lambda_1(\theta)\} + \Lambda_2(\theta), \\ p_{ij}(\theta) &= A_{ij} \exp\{K_{ij} \Lambda_1(\theta)\}, \\ p_{\beta 0}(\theta) &= B_\beta \exp\{M_\beta \Lambda_1(\theta) + V_2(\theta)\}, \end{aligned}\right\} \quad (\alpha, \beta, i, j = 1, 2, \ldots, s), \tag{3·4}$$

where as before $\Lambda_1(\theta) = \int k(\theta)\,d\theta$ and $\Lambda_2(\theta) = \int u(\theta)\,d\theta$, while $V_2(\theta) = \int v(\theta)\,d\theta$.

We now show that since $\sum_{j=0}^{s} p_{ij}(\theta) = 1$ for each $i = 1, \ldots, s$, all $p_{\beta 0}(\theta)$ $(\beta = 1, \ldots, s)$ must be identical. For consider the two rows i_1 and i_2; then clearly for these

$$\left\{1 - \sum_{j=1}^{s} p_{i_1 j}(\theta)\right\} \Big/ B_{i_1} \exp\{M_{i_1} \Lambda_1(\theta)\} = \left\{1 - \sum_{j=1}^{s} p_{i_2 j}(\theta)\right\} \Big/ B_{i_2} \exp\{M_{i_2} \Lambda_1(\theta)\} = \exp V_2(\theta)$$

for all values of θ. Equating the single positive terms in each expression, we obtain that $B_{i_1} = B_{i_2}$ and $M_{i_1} = M_{i_2}$, so that all

$$p_{\beta 0}(\theta) = 1 - \sum_{j=1}^{s} p_{\beta j}(\theta) = f_0(\theta) \quad (\beta = 1, \ldots, s) \tag{3·5}$$

are identical. It is convenient to write these as $f_0(\theta) = \exp V_2(\theta)$ where $B_\beta = 1$ and $M_\beta = 0$. Thus, exactly as for the p.r. chain, there exist only $r \leqslant s$ distinct exponents K_{ij} $(i, j = 1, \ldots, s)$ such that apart from coefficients A_{ij} the $p_{ij}(\theta)$ other than p_{i0} in each row of the stochastic matrix consist of the same $r \leqslant s$ distinct forms of type (3·4) in varying orders.

Lastly, the function G is given by

$$G(T, \theta; x, y) = k(\theta)\, g(T) + xu(\theta) + yv(\theta), \tag{3·6}$$

where

$$g(T) = \sum_{\alpha=1}^{s} x_\alpha K_\alpha + \sum_{i,j=1}^{s} n_{ij} K_{ij} \tag{3·7}$$

precisely as in (2·12).

4. Unbiasedness and minimum variance properties

In distributions admitting a sufficient statistic, the estimator is known to have properties of unbiasedness and minimum variance (Rao, 1949). For the p.r. Markov chain of §2, the distribution of the x_α and n_{ij} in a realization with a fixed number $n = \sum_{i,j=1}^{s} n_{ij}$ of transitions is

$$P(S) = C \exp\left\{ \sum_{\alpha=1}^{s} x_\alpha K_\alpha \theta + \Lambda_2(\theta) + \sum_{i,j=1}^{s} n_{ij} K_{ij} \theta + n\lambda_2(\theta) + \sum_{\alpha=1}^{s} x_\alpha \ln A_\alpha + \sum_{i,j=1}^{s} n_{ij} \ln A_{ij} \right\}, \tag{4·1}$$

where we have written θ for $\Lambda_1(\theta)$ for simplicity, and C is some constant.

Now carrying out the summation Σ' of $P(S)$ in (4·1) over all possible values of the n_{ij} starting with state α and such that $\sum_{i,j=1}^{s} n_{ij} = n$ remains fixed, and then over all states α, we obtain

$$\sum_{\alpha=1}^{s} \Sigma' P(S) = 1. \tag{4·2}$$

It follows readily on differentiating with respect to θ that

$$E\left(\sum_{\alpha=1}^{s} x_\alpha K_\alpha + \sum_{i,j=1}^{s} n_{ij} K_{ij} \right) = -\{n\lambda_2'(\theta) + \Lambda_2'(\theta)\}, \tag{4·3}$$

so that $g(T) = \sum_{\alpha=1}^{s} x_\alpha K_\alpha + \sum_{i,j=1}^{s} n_{ij} K_{ij}$ is an unbiased estimator of $-\{n\lambda_2'(\theta) + \Lambda_2'(\theta)\}$. A second differentiation of (4·2) with respect to θ gives the (minimum) variance of $g(T)$ as

$$\vartheta(g(T)) = -\{n\lambda_2''(\theta) + \Lambda_2''(\theta)\}. \tag{4·4}$$

In a similar manner, the distribution of the x_α, n_{ij} in the r. Markov chain of §3 for a realization with a random number $n+1$ of transitions, where $n = \sum_{i,j=1}^{s} n_{ij}$ is given by

$$P(S') = C \exp\left\{ \sum_{\alpha=1}^{s} x_\alpha K_\alpha \theta + \Lambda_2(\theta) + \sum_{i,j=1}^{s} n_{ij} K_{ij} \theta + V_2(\theta) + \sum_{\alpha=1}^{s} x_\alpha \ln A_\alpha + \sum_{i,j=1}^{s} n_{ij} \ln A_{ij} \right\}, \tag{4·5}$$

where $\Lambda_1(\theta)$ is again written θ for simplicity, and C is a constant.

Summing over all possible n_{ij} for a fixed value n, then over all $n = 0, 1, \ldots$, and finally over all α, we have that

$$\sum_{\alpha=1}^{s} \sum_{n=0}^{\infty} \Sigma' P(S') = 1, \tag{4·6}$$

and as before, differentiation with respect to θ gives

$$E\left(\sum_{\alpha=1}^{s} x_\alpha K_\alpha + \sum_{i,j=1}^{s} n_{ij} K_{ij} \right) = -\{\Lambda_2'(\theta) + V_2'(\theta)\}. \tag{4·7}$$

so that

$$g(T) = \sum_{\alpha=1}^{s} x_\alpha K_\alpha + \sum_{i,j=1}^{s} n_{ij} K_{ij}$$

is an unbiased estimator of $-\{\Lambda_2'(\theta) + V_2'(\theta)\}$. Its (minimum) variance is

$$\vartheta(g(T)) = -\{\Lambda_2''(\theta) + V_2''(\theta)\}. \tag{4·8}$$

5. Distributions of the sufficient estimators

Let $\phi(t)$ be the characteristic function (c.f.)

$$\phi(t) = E(e^{itg(T)}) \tag{5·1}$$

of the sufficient estimators

$$g(T) = \sum_{\alpha=1}^{s} x_\alpha K_\alpha + \sum_{i,j=1}^{s} n_{ij} K_{ij}$$

in both the cases considered. We continue to write θ for $\Lambda_1(\theta)$.

For the p.r. Markov chain of § 2, we have from (4·1) that

$$\begin{aligned}\phi(t) &= \sum_{\alpha}\sum{}' C \exp\{(\theta+it)(\sum_{\alpha} x_\alpha K_\alpha + \sum_{i,j} n_{ij}K_{ij}) + \Lambda_2(\theta) + n\lambda_2(\theta) + \sum_{\alpha} x_\alpha \ln A_\alpha + \sum_{i,j} n_{ij} \ln A_{ij}\} \\ &= \exp\{\Lambda_2(\theta) + n\lambda_2(\theta) - \Lambda_2(\theta+it) - n\lambda_2(\theta+it)\} \qquad (5\cdot2)\end{aligned}$$

as a result of (4·2). This may readily be rewritten as

$$\begin{aligned}\phi(t) &= \sum_{\alpha} A_\alpha \exp\{K_\alpha(\theta+it) + \Lambda_2(\theta)\}(\sum_{j} A_{rj} \exp\{K_{rj}(\theta+it) - \lambda_2(\theta)\})^n \\ &= \sum_{\alpha} \alpha_\alpha(\theta) \exp\{iK_\alpha t\}(\sum_{j} p_{rj}(\theta) \exp\{iK_{rj}t\})^n, \qquad (5\cdot3)\end{aligned}$$

a form similar to the c.f. for a multinomial distribution. The asymptotic normality of the estimator as $n \to \infty$ follows readily.

For the r. Markov chain of § 3, we have from (4·5) that

$$\begin{aligned}\phi(t) &= \sum_{\alpha}\sum_{n}\sum{}' C \exp\{(\theta+it)(\sum_{\alpha} x_\alpha K_\alpha + \sum_{i,j} n_{ij}K_{ij}) + \Lambda_2(\theta) + V_2(\theta) + \sum_{\alpha} x_\alpha \ln A_\alpha + \sum_{i,j} n_{ij} \ln A_{ij}\} \\ &= \exp\{\Lambda_2(\theta) + V_2(\theta) - \Lambda_2(\theta+it) - V_2(\theta+it)\} \qquad (5\cdot4)\end{aligned}$$

as a result of (4·6). This may be rewritten as

$$\phi(t) = f_0(\theta)\{f_0(\theta+it)\}^{-1} \sum_{\alpha} a_\alpha(\theta) \exp\{iK_\alpha t\}. \qquad (5\cdot5)$$

It is interesting to note that (5·5) depends on the initial probabilities and $f_0(\theta)$ only. If m distinct realizations are taken, the c.f. will be $\phi^m(t)$, and again as $m \to \infty$, the estimator is readily shown to be asymptotically normal (cf. Bhat, 1959).

We illustrate the previous results briefly in tl.e following examples.

Example 1. If in the p.r. Markov chain, the transition probabilities in a particular row i of the stochastic matrix are binomial, so that for $\theta < 0$

$$p_{ij}(\theta) = \binom{s}{j} e^{j\theta}(1-e^\theta)^{s-j} \quad (j = 1, 2, \ldots, s) \qquad (5\cdot6)$$

and the remaining rows are varying permutations of these, then

$$\begin{aligned}\phi(t) &= \sum_{r} a_r(\theta) \exp\{iK_r t\}\left(\sum_{j}\binom{s}{j} e^{j(\theta+it)}(1-e^\theta)^{s-j}\right)^n \\ &= \sum_{r} a_r(\theta) \exp\{iK_i t\}(e^{\theta+it} + (1-e^\theta))^m, \qquad (5\cdot7)\end{aligned}$$

so that the distribution of the estimator $g(T)$ is the convolution of the initial probability distribution with a binomial distribution.

Example 2. Let all the K_{ij} in the r. Markov chain be constants equal to unity (say), then

$$g(T) = \sum_{i=1}^{s} x_i K_i + n$$

is a sufficient estimator of θ. Now for any row $r = 1, \ldots, s$, we have that

$$\begin{aligned}f_0(\theta+it) &= 1 - \sum_{j=1}^{s} A_{rj} e^{(\theta+it)} \\ &= 1 - e^{it} \sum_{j=1}^{s} A_{ij} e^\theta = 1 - e^{it}(1-f_0(\theta)). \qquad (5\cdot8)\end{aligned}$$

It follows that (5) reduces to

$$\phi(t) = f_0(\theta)\{1 - e^{it}(1-f_0(\theta))\}^{-1} \sum_{r} a_r(\theta) \exp\{iK_r t\}, \qquad (5\cdot9)$$

so that the distribution of the sufficient estimator $g(T)$ is the convolution of the initial probability with a geometric distribution (cf. Bhat, 1959).

REFERENCES

BHAT, B. R. (1960). Some properties of the regular Markov chain. (To appear.)
GANI, J. (1955). Some theorems and sufficiency conditions for the maximum likelihood estimator of an unknown parameter in a simple Markov chain. *Biometrika*, **42**, 342–59.
GANI, J. (1956). Corrigenda. *Biometrika*, **43**, 497–8.
RAO, C. R. (1949). Sufficient statistics and minimum variance estimates. *Proc. Camb. Phil. Soc.* **45**, 213–8.

The regression of true value on estimated value

BY R. N. CURNOW

*A.R.C. Unit of Statistics, University of Aberdeen, Aberdeen, Scotland**

1. INTRODUCTION AND SUMMARY

Selection has often to be based on the estimated values of a certain character rather than on the true values of that character. Examples abound in plant and animal breeding work and also, for example, in selection based on estimated I.Q. scores. The effect of the selection on the mean estimated value can be readily calculated. The effect of the selection on the mean true value is clearly of much greater importance and its calculation requires a knowledge of the regression of true value on estimated value.

We shall assume that the true value, x, is estimated by y, where

$$y = x + w,$$

with w normally distributed independently of x with zero mean and variance ϵ^2. We shall further assume that x is distributed continuously over the population subjected to selection with frequency function $f(x)$. All the integrations with respect to x can be replaced by summations to show that the results also apply when x has a discrete distribution. The joint frequency function of x and y is

$$f(x, y) = \frac{1}{\epsilon} f(x)\, \phi\left(\frac{y-x}{\epsilon}\right),$$

where

$$\phi(z) = \frac{1}{(2\pi)^{\frac{1}{2}}} \exp\{-\tfrac{1}{2}z^2\}.$$

Therefore the regression of x on y is

$$h(y) = E(x|y) = \frac{\frac{1}{\epsilon}\int_{-\infty}^{+\infty} x f(x)\, \phi\left(\frac{y-x}{\epsilon}\right) dx}{\frac{1}{\epsilon}\int_{-\infty}^{+\infty} f(x)\, \phi\left(\frac{y-x}{\epsilon}\right) dx}.$$

If selection improves the average value of y by an amount Δy, then the expected value of the corresponding increase in the average value of x will be

$$\Delta x = h(y + \Delta y) - h(y).$$

If x is normally distributed with mean μ and variance σ^2, the function $h(y)$ is known to be

$$h(y) = \mu + \frac{\sigma^2}{\sigma^2 + \epsilon^2}(y - \mu).$$

Therefore,

$$\Delta x = \frac{\sigma^2}{\sigma^2 + \epsilon^2} \Delta y.$$

Lindley (1947) showed that the necessary and sufficient condition for x to regress linearly on y is that the cumulant generating function of the distribution of x be a multiple of the cumulant generating function of the distribution of w. As an example, he showed that the regression is linear if x and w have independent χ^2 distributions. We shall show that when w is normally distributed, the form of the

* This work was completed while the author was visiting Iowa State University on a Harkness Fellowship of the Commonwealth Fund.

regression always completely determines the distribution of x. Lindley's theorem and this last result have a special case in common—when w is normally distributed linearity of regression implies a normal distribution of x.

A linear regression of true value on estimated value is often assumed in the calculation of the consequences of selection. Few writers known to the author have stated that, if the errors are normally distributed, this involves an assumption of normality for the distribution of true values. Departures from normality in the error distribution may increase or decrease the non-linearity of the regression depending on the distribution of x. For example, if x and w both happen to have χ^2 distributions the regression will be exactly linear. Clearly, small increases in the expected mean value of x will be approximately linearly related to small increases in the mean value of y whatever the distribution of x. However, even in this case, the actual value of the regression coefficient, i.e. the rate of increase of x with y, will be a function of y and only exceptionally the same as that calculated assuming normality. The functions $h(y)$ will be given for certain non-normal distributions of x.

2. Derivation of $f(x)$ given $h(y)$

Writing
$$u(y) = \frac{1}{\epsilon}\int_{-\infty}^{+\infty} x f(x)\,\phi\left(\frac{y-x}{\epsilon}\right)dx$$

and
$$v(y) = \frac{1}{\epsilon}\int_{-\infty}^{+\infty} f(x)\,\phi\left(\frac{y-x}{\epsilon}\right)dx,$$

we have that
$$h(y) = \frac{u(y)}{v(y)}.$$

Now, differentiating $v(y)$,
$$\frac{dv(y)}{dy} = -\frac{1}{\epsilon^2}\int_{-\infty}^{+\infty}\frac{(y-x)}{\epsilon} f(x)\,\phi\left(\frac{y-x}{\epsilon}\right)dx$$
$$= \frac{1}{\epsilon^2}[u(y) - yv(y)]$$
$$= \frac{v(y)}{\epsilon^2}[h(y) - y].$$

The solution of this equation is
$$v(y) = c\exp\left\{\frac{H(y) - \frac{1}{2}y^2}{\epsilon^2}\right\},$$

where $H(y) = \int^y h(z)\,dz$ and c is an arbitrary constant. Now $v(y)$ is the frequency function of the distribution of y and therefore the characteristic function is

$$\phi_y(t) = c\int_{-\infty}^{+\infty}\exp\left\{\frac{H(y) - \frac{1}{2}y^2 + iyt\epsilon^2}{\epsilon^2}\right\}dy.$$

Finally, we know that x and w are independently distributed and therefore that the characteristic function of the distribution of x is

$$\phi_x(t) = \phi_y(t)/\phi_w(t)$$
$$= c\exp\left\{\frac{\epsilon^2 t^2}{2}\right\}\int_{-\infty}^{+\infty}\exp\left\{\frac{H(y) - \frac{1}{2}y^2 + iyt\epsilon^2}{\epsilon^2}\right\}dy.$$

The constant c is chosen so that $\phi_x(0) = 1$. Since a frequency function is uniquely determined by its characteristic function, this completes the proof that the regression of x on y completely determines the distribution of x. The converse, that the distribution of x completely determines the regression of x on y, is obviously true. Since $\phi_x(t)$ has to satisfy certain conditions to be a characteristic function, not every function $h(y)$ is a possible regression function.

Since
$$\frac{du(y)}{dy} = \frac{1}{\epsilon^2}\left[\frac{1}{\epsilon}\int_{-\infty}^{+\infty} x^2 f(x)\,\phi\left(\frac{y-x}{\epsilon}\right)dx - yu(y)\right]$$

and
$$\frac{dv(y)}{dy} = \frac{1}{\epsilon^2}[u(y) - yv(y)],$$

$$\frac{dh(y)}{dy} = \frac{1}{v(y)}\frac{du(y)}{dy} - \frac{u(y)}{v^2(y)}\frac{dv(y)}{dy}$$

$$= \frac{1}{\epsilon^2}\left[\frac{1}{\epsilon v(y)}\int_{-\infty}^{+\infty} x^2 f(x)\,\phi\left(\frac{y-x}{\epsilon}\right)dx - h^2(y)\right]$$

$$= \frac{1}{\epsilon^2}[E(x^2|y) - [E(x|y)]^2]$$

$$= \frac{\text{variance of } x \text{ for given } y}{\epsilon^2}.$$

Therefore, we have also shown that, given the mean of x, the distribution of x is completely determined once the dependence of the variance of x for given y on y is known. More importantly, perhaps, this shows that if selection increases y by a small amount from y_0 to $y_0 + \Delta y_0$, then the corresponding increase in x will be approximately

$$\Delta x_0 = \frac{\text{variance of } x \text{ when } y = y_0}{\epsilon^2}\Delta y_0.$$

3. Some special cases

If $h(y)$ is linear, then x is normally distributed. This follows immediately by writing

$$h(y) = a + b(y-a)$$

and, therefore,

$$H(y) = ay + \tfrac{1}{2}b(y-a)^2.$$

Substituting in the expression for $\phi_x(t)$ and integrating,

$$\phi_x(t) = \exp\left\{ait - \frac{\epsilon^2 bt^2}{2(1-b)}\right\}.$$

Therefore x is normally distributed with mean a and variance $b\epsilon^2/(1-b)$. In the previous notation $a = \mu$ and $b = \sigma^2/(\sigma^2+\epsilon^2)$.

The functions $h(y)$ can be written in terms of well-tabulated functions for certain distributions of x. For example, if x has a rectangular distribution between $x = 0$ and $x = 1$,

$$h(y) = y - \epsilon\frac{\phi\left(\frac{y-1}{\epsilon}\right) - \phi\left(\frac{y}{\epsilon}\right)}{\Phi\left(-\frac{y-1}{\epsilon}\right) - \Phi\left(-\frac{y}{\epsilon}\right)},$$

where

$$\Phi(z) = \int_{-\infty}^{z}\phi(x)\,dx.$$

If x has a χ^2 distribution with ν degrees of freedom,

$$h(y) = \frac{\nu\epsilon}{2}\frac{Hh_{\frac{1}{2}\nu}\left(\frac{\epsilon}{2} - \frac{y}{\epsilon}\right)}{Hh_{\frac{1}{2}\nu-1}\left(\frac{\epsilon}{2} - \frac{y}{\epsilon}\right)},$$

where Hh_n are the Hermitian Probability Functions tabulated for integral values of n in the *British Association Mathematical Tables* (1951). If x is distributed as $\sqrt{\chi^2}$, where χ^2 has ν degrees of freedom, then

$$h(y) = \frac{\nu\epsilon}{\sqrt{(1+\epsilon^2)}}\frac{Hh_\nu\left(-\frac{y}{\epsilon\sqrt{(1+\epsilon^2)}}\right)}{Hh_{\nu-1}\left(-\frac{y}{\epsilon\sqrt{(1+\epsilon^2)}}\right)}.$$

With appropriate changes in the scale and origin of y and $h(y)$, the function $h(y)$ for χ^2 with 2ν degrees of freedom is the same as the function $h(y)$ for $\sqrt{\chi^2}$ where χ^2 has ν degrees of freedom. This does not contradict the theorem above that the regression uniquely determines the distribution of x. For distributions such as these in which the range of possible values of x does not extend from $-\infty$ to $+\infty$, a linear regression of x on y over the whole range of values of y is obviously impossible.

The magnitude of the deviations from linearity will clearly depend on the ratio, ϵ^2/σ^2, of error variance to population variance, the extent of the departure from normality in the distribution of x and the intensity of the selection. When selection is aimed at increasing the value of a character that has a definite upper limit, and that upper limit has nearly been reached, then the regression is likely to be far from linear. This will occur, for example, in selection from rectangular and 'reversed' χ^2 and χ distributions. The regression function for the 'reversed' distribution [i.e. the distribution with frequency function equal to $f(-x)$ instead of $f(x)$ (see Curnow, 1958)] is obtained from $h(y)$, the regression function for the original distribution, as $-h(-y)$.

I am grateful to Dr D. J. Finney and Prof. O. Kempthorne for suggesting improvements in the presentation of this paper.

References

British Association Mathematical Tables (1951). Vol. I. 3rd edition. Cambridge University Press.

Curnow, R. N. (1958). The consequences of errors of measurement for selection from certain non-normal distributions. *Proceedings of the 31st Session of the International Statistical Institute.*

Lindley, D. V. (1947). Regression lines and the linear functional relationship. *J. R. Statist. Soc.* Suppl. **9**, 218–44.

The use of integral transforms to determine expansions of distribution functions

By W. M. HARPER and J. A. MACDONALD

Australian Defence Scientific Service, Defence Standards Laboratories, Maribyrnong, Victoria, Australia

1. Introduction

For many variates, the evaluation of probability points is extremely tedious. In the case of the quadratic form

$$x = \Sigma\lambda_j u_j^2,$$

for example, where the u_j are independent standard normal variates and the λ_j are constants, many methods which might be used to determine the distribution have been devised including among others those given by Robbins & Pitman (1949), Grad & Solomon (1955), Gurland (1955, 1957) and Daniels (1956). The usefulness of these methods, however, is markedly affected by the values and multiplicities of the λ_j.

If the characteristic function of the variate can be obtained, some integral transforms of the distribution function are easily derived. The transforms considered here are the Fourier and Laplace transforms, the latter being more convenient for variates which are always of the same sign.

Under certain conditions it is then possible to obtain expansions from which probability points in both tails of the probability distribution curve can be evaluated. The procedure appears to be useful for dealing with some quadratic forms where other methods are unsuitable or inconvenient.

Although in this paper emphasis is placed on statistical distributions, the expansions have other applications in analysis (see, for example, Watson (1945), Jeffreys & Jeffreys (1946), Carslaw & Jaeger (1948)).

2. Expansions for the distribution function

In order to ensure convergence at infinity of the Fourier integral used in the following analysis, a 'modified distribution function' is defined by

$$R(x) = H(x) - \int_{-\infty}^{x} f(y)\,dy,$$

where $H(x)$ is Heaviside's unit function and $f(x)$ the probability density function of the variate x, i.e.

$$H(x) = 0, \quad R(x) = -P(x) \qquad (x < 0),$$
$$H(x) = 1, \quad R(x) = 1 - P(x) = Q(x) \quad (x > 0),$$

where $P(x)$ is the distribution function.

In terms of the characteristic function $\phi(t)$, the Fourier transform $\rho(t)$ of $R(x)$ is given by

$$\begin{aligned}\rho(t) &= \int_{-\infty}^{\infty} e^{itx} R(x)\,dx \\ &= (it)^{-1}\{\phi(t)-1\},\end{aligned}$$

by partial integration. The Fourier inversion formula becomes

$$R(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-ixt}\rho(t)\,dt. \tag{1}$$

As remarked above, for positive variates it is more convenient to use the Laplace transformation; no modification of the distribution function is necessary in this case as convergence at $-\infty$ is not in question. The Laplace transform $\bar{P}(p)$ of the distribution function is related to the characteristic function by

$$\begin{aligned}\bar{P}(p) &= \int_0^{\infty} e^{-px}\,dx \int_0^x f(y)\,dy \\ &= \frac{1}{p}\int_0^{\infty} e^{-py} f(y)\,dy \\ &= p^{-1}\phi(ip).\end{aligned}$$

The transform $\bar{Q}(p)$ of $Q(x)$ is given by

$$\begin{aligned}\bar{Q}(p) &= p^{-1} - \bar{P}(p) \\ &= \rho(ip),\end{aligned}$$

where the function ρ is as defined above. The standard inversion formula becomes

$$Q(x) = \frac{1}{2\pi i}\int_{\gamma-i\infty}^{\gamma+i\infty} e^{xp}\rho(ip)\,dp, \tag{2}$$

where γ is an arbitrary positive number (there being no singularity of $\bar{Q}(p)$ with positive real part).

For evaluation of probability points in the lower tail of the probability distribution curve of positive variates, an extension of a known result of Laplace transformation theory can be used. It can be shown that it $\bar{P}(p)$ can be expanded as the series

$$\bar{P}(p) = \sum_{j=0}^{\infty} a_j (p-\alpha)^{\mu_j} \quad (0 > \mu_1 > \mu_2 > \ldots),$$

assumed uniformly convergent for $|p-\alpha|$ sufficiently large, then

$$P(x) = e^{\alpha x}\sum_{j=0}^{\infty} a_j\{\Gamma(-\mu_j)\}^{-1} x^{-\mu_j-1};$$

the case $\alpha = 0$ is proved by, for example, Carslaw & Jaeger (1948). The most convenient value of α will depend on the variate; in general a non-zero value will give the most rapidly convergent series.

For evaluation of probability points for which $|x|$ is large, asymptotic expansions can be obtained if $\rho(t)$ can be written in the form

$$\rho(t) = \rho_1(t) + \rho_2(t),$$

where

$$\rho_1(t) = \sum_{j=0}^{n} b_j(\omega \pm it)^{v_j},$$

$$\rho_2(t) = \sum_{j=n+1}^{\infty} b_j(\omega \pm it)^{v_j},$$

where

$$\mathscr{R}(\omega) > 0, \quad v_0 < v_1 < \ldots, \quad v_{n+1} > -1$$

and the series for $\rho_2(t)$ are assumed uniformly convergent for $|\omega \pm it|$ sufficiently small.

In the inversion formula (1), the line of integration may be replaced by any other contour C such that there are no singularities of $\rho(t)$ between C and the real axis, and the integrand is sufficiently small at infinity. (An analogous procedure can be followed for the inversion formula (2) in the case of positive variates.) Thus for large negative x the plus sign is to be used, and for large positive x the minus sign is to be used; in both cases the expansion is about the singularity nearest the real axis. If there should be two or more singularities with the same imaginary part there will be a contribution to the expansion from each of them.

The contribution of $\rho_1(t)$ is

$$R_1(x) = \sum_{j=0}^{n} \frac{b_j}{2\pi} \int_C e^{\pm i|x|t} (\omega \pm it)^{v_j}\, dt;$$

here the integrand has at most one singularity and the contour C may be replaced by a contour surrounding this point (positively or negatively in the respective cases) as in the derivation of Hankel's integral for the gamma-function (Whittaker & Watson (1946)). The substitution

$$\omega \pm it = -t'/|x|$$

then gives

$$R_1(x) = \sum_{j=0}^{n} \frac{b_j}{2\pi} \int_{\infty}^{(0+)} i\, e^{-\omega|x|}\, |x|^{-v_j-1} e^{-t'} (-t')^{v_j}\, dt'$$

$$= e^{-\omega|x|} \sum_{j=0}^{n} b_j \{\Gamma(-v_j)\}^{-1} |x|^{-v_j-1}.$$

The proof of a similar expansion is given by Carslaw & Jaeger (1948) but the series is given in a slightly different form and the unnecessary restriction that $v_0 \geqslant -1$ is imposed.

The contribution of $\rho_2(t)$, denoted by $R_2(x)$, may be estimated as in Carslaw & Jaeger (1948) with the appropriate change of axis; the result is that

$$R_2(x) = o[\exp\{-\mathscr{R}(\omega)\,|x|\}\,|x|^{-v_j-1}].$$

Hence we have an asymptotic expansion

$$R(x) \sim e^{-\omega|x|} \sum_{j=0}^{\infty} b_j \{\Gamma(-v_j)\}^{-1} |x|^{-v_j-1}$$

from which probability points in both tails of the probability distribution curve can be evaluated.

3. Application

Chernoff & Lehmann (1954) have shown that in making a χ^2 goodness-of-fit test, estimation of the distribution parameters from the sample values instead of the cell frequencies leads to a statistic which is not asymptotically distributed as chi-square but as

$$\sum_{j=1}^{k-s-1} u_j^2 + \sum_{j=k-s}^{k-1} \lambda_j u_j^2,$$

where k is the number of cells, s is the number of parameters to be estimated, the u_j are independent standard normal variates and the λ_j are constants, When, for example, in fitting a distribution the mean and variance are estimated by the sample mean and variance, the statistic obtained has the limiting distribution

$$X^2 = \chi^2_{k-3} + \lambda_1 u_1^2 + \lambda_2 u_2^2,$$

where χ^2_{k-3} is a chi-square variate with $k-3$ degrees of freedom, u_1, u_2 are standard normal variates independent of each other and of χ^2_{k-3} and λ_1, λ_2 are constants less than unity. Watson (1957) suggests that in practical applications, the probability that X^2 exceeds x should be calculated where x is the upper 5 % point of χ^2_{k-3}. For a normal distribution and the case $k = 10$, for example, the relevant value of x is 14·067 and Watson has shown the values of λ_1 and λ_2 to be respectively 0·042 and 0·297.

For such problems the method of the present paper is well suited; for

$$\bar{Q}(p) = p^{-1} - p^{-1}(1+2p)^{-\frac{1}{2}(k-3)}\,(1+2\lambda_1 p)^{-\frac{1}{2}}\,(1+2\lambda_2 p)^{-\frac{1}{2}},$$

and expansion of $\bar{Q}(p)$ about the singularity at $-\frac{1}{2}$ leads to the asymptotic expansion

$$Q(x) \sim \frac{e^{-\frac{1}{2}x}\, x^{\frac{1}{2}(k-5)}}{2^{\frac{1}{2}(k-5)}\Gamma\{\frac{1}{2}(k-3)\}\,(1-\lambda_1)^{\frac{1}{2}}\,(1-\lambda_2)^{\frac{1}{2}}} \left\{1 + \frac{c_1}{x} + \frac{c_2}{x^2} + \ldots\right\},$$

where the first two coefficients in the series are given by

$$c_1 = (k-5)\,\{1 - \tfrac{1}{4}(\alpha_1+\alpha_2)\},$$

$$c_2 = (k-5)\,(k-7)\,\{1 - \tfrac{1}{4}(\alpha_1+\alpha_2) + \tfrac{3}{32}(\alpha_1^2+\alpha_2^2) + \tfrac{1}{16}\alpha_1\alpha_2\},$$

$$\alpha_1 = 2\lambda_1/(1-\lambda_1), \quad \alpha_2 = 2\lambda_2/(1-\lambda_2).$$

These terms are sufficient to give the result of 0·057 correct to three decimal places for the numerical values quoted above, while two more terms would give $Q(x)$ correct to five places of decimals if required.

Other methods seem more awkward to use in this application. The method of mixtures developed by Robbins & Pitman (1949), for example, yields a series which converges much too slowly to be of any use,

while other methods appear to be equally laborious. In general the same considerations apply to the more general quadratic form

$$X^2 = \sum_{j=1}^{n} \lambda_j \chi^2_{kj}$$

whenever one λ_j is much greater than the others.

The authors are indebted to Dr G. S. Watson of the Australian National University for suggesting extension of the work to cover indefinite variates and for other helpful suggestions and comments. The paper is published with the permission of the Chief Scientist, Australian Defence Scientific Service, Department of Supply, Melbourne, Australia.

References

Carslaw, H. S. & Jaeger, J. C. (1948). *Operational Methods in Applied Mathematics*, 2nd ed. Oxford University Press.

Chernoff, H. & Lehmann, E. L. (1954). The use of the maximum likelihood estimates in χ^2 tests for goodness of fit. *Ann. Math. Statist.* **25**, 579–86.

Daniels, H. E. (1956). The approximate distribution of serial correlation coefficients. *Biometrika*, **43**, 169–85.

Grad, A. & Solomon, H. (1955). Distribution of quadratic forms and some applications. *Ann. Math. Statist.* **26**, 464–77.

Gurland, J. (1955). Distribution of definite and of indefinite quadratic forms. *Ann. Math. Statist.* **26**, 122–7.

Gurland, J. (1957). Quadratic forms in normally distributed random variables. *Sankhyā*, **17**, 37–50.

Jeffreys, H. & Jeffreys, B. S. (1946). *Methods of Mathematical Physics*. Cambridge University Press.

Robbins, H. & Pitman, E. J. G. (1949). Application of the method of mixtures to quadratic forms in normal variates. *Ann. Math. Statist.* **20**, 552–60.

Watson, G. N. (1945). *Theory of Bessel Functions*, 2nd ed. Cambridge University Press.

Watson, G. S. (1957). The χ^2 goodness-of-fit test for normal distributions. *Biometrika*, **44**, 336–48.

Whittaker, E. T. & Watson, G. N. (1946). *A Course of Modern Analysis*, 4th ed. reprinted. Cambridge University Press.

Some applications of two approximations to the multinomial distribution

By N. L. Johnson and D. H. Young

University College London

1. Introduction

The derivation of exact distributions of statistics based on multinomial variables is not simple, except in a few special cases. If $n_1, n_2, \ldots, n_k$ have the joint multinomial distribution

$$P(n_1, n_2, \ldots, n_k) = N! \prod_{j=1}^{k} (\pi_j^{n_j}/n_j!), \tag{1}$$

where $\sum_{j=1}^{k} n_j = N$, $\pi_j \geqslant 0$ and $\sum_{j=1}^{k} \pi_j = 1$, then the distributions of $n_j, n_j + n_{j'}$, etc., are simply obtained as binomials with parameters N, π_j; $N, \pi_j + \pi_{j'}$, etc. But when studying the distribution of statistics such as $\chi^2 = \sum_{j=1}^{k} (n_j - N\pi_j)^2 (N\pi_j)^{-1}$, etc., some form of approximation is useful.

In this paper two approximations to (1) are used in deriving approximate significance limits of two criteria which might be used as alternatives to the standard χ^2-test of the hypothesis

$$\pi_1 = \pi_2 = \ldots = \pi_k = k^{-1}.$$

The suggested criteria are $R = \dfrac{\min n_j}{\max n_j}$ and $W = \dfrac{\max n_j - \min n_j}{N}$,

which are simpler to calculate than is χ^2 (though neither R nor W can usefully be used if the π_j's are not equal). In order to use these criteria we need tables of lower percentage points of the distribution of R and tables of upper percentage points of the distribution of W, when $\pi_1 = \pi_2 = \ldots = \pi_k = k^{-1}$. A number of approximations to these values are considered.

Also in a preliminary investigation, three approximations to the distribution of the maximum relative frequency, $\max(n_j/N)$, will be considered.

2. An approximation to the multinomial distribution

Putting $p_j = n_j/N$ and treating the p_j's as continuous variables, we may use the approximate joint probability density function

$$g(p_1, p_2, \ldots, p_k) = \Gamma(N-1) \prod_{j=1}^{k} \left(\frac{p_j^{(N-1)\pi_j - 1}}{\Gamma[(N-1)\pi_j]} \right) \quad \left(0 \leqslant p_j, \sum_{j=1}^{k} p_j = 1 \right). \tag{2}$$

$g(p_1, p_2, \ldots, p_k)$ is to be interpreted as giving the joint distribution of any $k-1$ of the p's; this is obtained by replacing the omitted p_l by $1 - \sum_{j \neq l}^{k} p_j$.

This approximate distribution gives correct first and second moments and product moments for the p's. Also it gives similar approximate distributions for subsets of the p's and for quantities such as $p_{i_1} + p_{i_2} + \ldots + p_{i_s}$ $(i_1 \neq i_2 \neq \ldots \neq i_s)$. Detailed examination of some properties of this approximation has been published in another paper (Johnson, 1960).

It is useful to note that, from the point of view of distribution theory, we can regard the approximate distribution of the p's as that of the ratios

$$\frac{\chi_1^2}{\sum_j \chi_j^2}, \quad \frac{\chi_2^2}{\sum_j \chi_j^2}, \ldots, \quad \frac{\chi_k^2}{\sum_j \chi_j^2}, \tag{3}$$

where $\chi_1^2, \chi_2^2, \ldots, \chi_k^2$ are independent χ^2's with $2(N-1)\pi_1, 2(N-1)\pi_2, \ldots, 2(N-1)\pi_k$ degrees of freedom, respectively.

3. An alternative approximation

An alternative approximation to the joint distribution of $n_1, n_2, \ldots, n_k$, which may be used if

$$\pi_1 = \pi_2 = \ldots = \pi_k = k^{-1},$$

may be obtained as follows.

Let us consider the standardized multinomial variables $w_i = (n_i - N\pi_i)/[N\pi_i(1-\pi_i)]^{\frac{1}{2}}$. It can be shown that as N increases the joint distribution of $w_1, w_2, \ldots, w_k$ tends to a degenerate multinormal distribution with zero means, unit variances and covariance between w_i and w_j equal to $-(\pi_i \pi_j)^{\frac{1}{2}}/[(1-\pi_i)(1-\pi_j)]^{\frac{1}{2}}$. If $\pi_1 = \pi_2 = \ldots = \pi_k = k^{-1}$ each covariance is equal to $-(k-1)^{-1}$.

Consider now a set of k *independent* unit normal variables $x_1, x_2, \ldots, x_k$. Then it can straightforwardly be shown that the standardized deviates from the sample mean

$$\left(\frac{k}{k-1}\right)^{\frac{1}{2}} (x_1 - \bar{x}), \left(\frac{k}{k-1}\right)^{\frac{1}{2}} (x_2 - \bar{x}), \ldots, \left(\frac{k}{k-1}\right)^{\frac{1}{2}} (x_k - \bar{x}) \tag{4}$$

are jointly distributed multinormally with zero means, unit variances and equal covariances $-(k-1)^{-1}$. Hence, when $\pi_1 = \pi_2 \ldots = \pi_k = k^{-1}$ the approximate multinormal distribution of the w's may be considered as the joint distribution of the standardized deviates given in (4).

Several properties of this approximation to the joint distribution of the w's have been considered by Young (1960).

4. Approximate upper percentage points of the distribution of max n_j

Using the approximation developed in § 2, we see that the distribution of $\max(n_j/N)$ may be approximated by that of Cochran's criterion

$$\max s_j^2 \Big/ \sum_{j=1}^{k} s_j^2,$$

where the s_j^2's are independent χ^2 variables, each with $2(N-1)k^{-1}$ degrees of freedom (assuming $\pi_1 = \pi_2 = \ldots \pi_k = k^{-1}$).

Tables of upper significance limits of this criterion are given in Eisenhart, Hastay & Wallis (1947, Chapter 15). Interpolating in these tables we find the values shown in columns (*a*) of Table 1.

Alternatively, if we use the approximation given in § 3, we see that the distribution of $\{(k-1)/k\}^{\frac{1}{2}} \max w_j$ may be approximated by that of $\max x_j - \bar{x}$, the extreme positive deviate from the sample mean in a sample of k independent unit normal variables. Since

$$\left(\frac{k-1}{k}\right)^{\frac{1}{2}} \max w_j = \left(\frac{k}{N}\right)^{\frac{1}{2}} \left(\max n_j - \frac{N}{k}\right)$$

it follows that the upper 100α % point of the distribution of $X = \max (n_j/N)$ is approximately given by

$$X_\alpha = \frac{1}{k} + \frac{u_\alpha}{(Nk)^{\frac{1}{2}}} + \frac{1}{2N}, \tag{5}$$

where u_α is the upper 100α % point of the distribution of $u = \max x_j - \bar{x}$.

Tables of u_α are given by Nair (1948), and these have been used to give the values of X_α shown in columns (*b*) of Table 1.

Table 1. *Approximate upper significance limits for* $\max (n_j/N)$ *when* $\pi_1 = \pi_2 = \ldots = \pi_k = k^{-1}$

	$\alpha = 0{\cdot}01$			$\alpha = 0{\cdot}05$		
k	(*a*)	(*b*)	(*c*)	(*a*)	(*b*)	(*c*)
			$N = 25$			
2	0·746	0·777	0·80	0·693	0·716	0·72
3	·606	·609	·64	·547	·554	·56
4	·532	·513	·56	·471	·464	·48
5	·476	·450	·48	·416	·406	·44
6	·440	·406	·44	·382	·365	·40
7	·414	·372	·44	·356	·334	·40
8	·393	·342	·40	·336	·310	·36
			$N = 50$			
2	0·678	0·692	0·70	0·637	0·649	0·66
3	·524	·524	·54	·482	·485	·50
4	·440	·432	·46	·399	·397	·40
5	·388	·373	·40	·348	·342	·36
6	·351	·331	·36	·312	·303	·32
7	·324	·300	·32	·286	·274	·30
8	·304	·276	·30	·267	·252	·28
			$N = 100$			
2	0·627	0·634	0·64	0·598	0·603	0·61
3	·469	·466	·47	·438	·439	·44
4	·384	·377	·38	·354	·352	·36
5	·329	·320	·33	·303	·298	·30
6	·292	·281	·28	·266	·261	·26
7	·266	·252	·26	·241	·234	·24
8	·246	·230	·23	·221	·213	·22

(*a*) based on Cochran's criterion; (*b*) using equation (5) and Nair's table; (*c*) using Young's approximation, with (6) and (7).

The factor $(2N)^{-1}$ which appears in (5) comes from a continuity correction, since possible values of $X = \max (n_j/N)$ are at intervals of N^{-1}.

Columns (*c*) of Table 1 show values obtained from further approximate formulae (Young, 1930)

$$kI_{k-1}[(NX_\alpha, N(1-X_\alpha)+1] \leqslant \alpha, \tag{6}$$

$$\Phi\left[\frac{N(kX_\alpha - 1) - \frac{1}{2}k}{\{N(k-1)\}^{\frac{1}{2}}}\right] \geqslant 1 - \frac{\alpha}{k}, \tag{7}$$

where
$$I_p(a, n-a+1) = 1 - \sum_{r=0}^{a-1} \binom{n}{r} p^r (1-p)^{n-r},$$

$$\Phi(x) = \frac{1}{\sqrt{(2\pi)}} \int_{-\infty}^{x} e^{-\frac{1}{2}t^2}\, dt,$$

and X_α is the least possible value of X satisfying (6) or (7).

Approximation (7) is useful when N is large and was used to obtain the significance limits in Table 1 (columns (*c*)) when $N = 100$; formula (6) was used for $N = 25, 50$. It can be shown that the maximum error in significance level as computed by the left-hand side of (6) is less than $\frac{1}{2}\alpha^2(1-k^{-1})$. In fact, all the values shown in columns (*c*) of Table 1 for $N = 25, 50$ are equal to the least X_α such that $\Pr\{X \geqslant X_\alpha\} \leqslant \alpha$.

It should be noted that, in the third approximation (columns (*c*)), values are given for max (n_j/N) corresponding only to integer values of n_j. This means that for $N = 25$, possible values are at intervals of 0·04, for $N = 50$ at intervals of 0·02, and for $N = 100$ at intervals of 0·01. In columns (*a*) and (*b*) values are given formally to three decimal places, but in fact they would give practical procedures identical with those given by the next higher (or equal) possible column (*c*) value.

Remembering this, we see that, for $k = 2$, approximation (5) is slightly better than the $\max s_j^2 \Big/ \sum_{j=1}^{k} s_j^2$ approximation. For $k = 3$, the three values given in table 1 would give identical practical procedures in every case. For values of $k \geqslant 4$, the $\max s_j^2 \Big/ \sum_{j=1}^{k} s_j^2$ approximation is more accurate than (5), especially for the smaller sample sizes; for $N = 25, 50$, the differences between the column (*b*) and column (*c*) values become more pronounced as k increases.

5. Approximate lower percentage points of the distribution of $R = \min n_j/\max n_j$

Noting that
$$R = \frac{\min n_j}{\max n_j} = \frac{\min p_j}{\max p_j},$$

§ 2 indicates that this ratio should be distributed approximately as $\min \chi_j^2/\max \chi_j^2$. If $\pi_j = k^{-1}$ for $j = 1, 2, \ldots, k$ then each χ^2 has the same number of degrees of freedom, $2(N-1)/k$. Table 31 of Pearson & Hartley (1954) gives significance limits for this statistic; from this table approximate significance limits for $\min n_j/\max n_j$ can be obtained by interpolation. Some further approximations can be obtained as follows.

If χ^2 has ν degrees of freedom, then $\log_e \chi^2$ is distributed approximately normally, with expected value $\Psi(\frac{1}{2}\nu) + \log_e 2$ and standard deviation $[\Psi^{(1)}(\frac{1}{2}\nu)]$, where

$$\Psi(x) = \frac{d}{dx}[\log_e \Gamma(x)] \quad \text{and} \quad \Psi^{(1)}(x) = \frac{d^2}{dx^2}[\log_e \Gamma(x)]$$

are the digamma and trigamma functions respectively (Bartlett & Kendall, 1946). Since $-\log_e R$ is approximately distributed as the range of logarithms of k independent χ^2's each with $2(N-1)/k$ degrees of freedom, the *lower* 100α % point of the distribution of R will be approximately equal to

$$\exp\left(-w_\alpha \sqrt{[\Psi^{(1)}\{(N-1)/k\}]}\right), \tag{8}$$

where w_α is the *upper* 100α % point of the distribution of range of k independent unit normal variables.

A further approximation, viz.
$$\Psi^{(1)}(x) \fallingdotseq (x-\tfrac{1}{2})^{-1}$$
which is quite good for $x \geqslant 2$, converts (8) to the simpler form

$$\exp\left(-w_\alpha \left(\frac{N-1}{k} - \tfrac{1}{2}\right)^{-\frac{1}{2}}\right). \tag{9}$$

Table 2 gives some approximate lower 5 and 1 % significance limits for R, using

(i) Table 31 of Pearson & Hartley (1954);
(ii) approximation (8);
(iii) approximation (9);
(iv) an approximation (to be described in § 6) which is obtained, using the approximate distribution of the standardized multinomial variables given in § 3.

The accuracy of the approximation $\Psi^{(1)}(x) \fallingdotseq (x-\frac{1}{2})^{-1}$ is so good that (ii) and (iii) agree to two places of decimals, even when x is as small as 2. Hence only one figure is shown for these two approximations.

Table 2. *Approximate lower significance limits for* $R = \min n_j / \max n_j$

(Actual significance levels in parentheses.)

	5% points			1% points		
k	(i)	(ii), (iii)	(iv)	(i)	(ii), (iii)	(iv)
			$N = 25$			
2	0·44 (0·0433) (0·1078*)	0·44 (0·0433) (0·1078*)	0·46 (0·0433) (0·1078*)	0·34 (0·0146) (0·0041*)	0·34 (0·0146) (0·0041*)	0·36 (0·0146) (0·0041*)
3	0·29 (0·0620) (0·0516*)	0·30 (0·0620) (0·0516*)	0·32 (0·0802) (0·0620*)	0·22 (0·0203) (0·0151*)	0·22 (0·0203) (0·0151*)	0·24 (0·0286) (0·0203*)
4	0·21	0·21	0·23	0·14	0·15	0·17
5	0·15	0·16	0·18	0·10	0·11	0·13
6	0·11	0·12	0·14	0·07	0·08	0·10
			$N = 50$			
2	0·57 (0·0649) (0·0328*)	0·57 (0·0649) (0·0328*)	0·58 (0·0649) (0·0328*)	0·47 (0·0153) (0·0066*)	0·48 (0·0153) (0·0066*)	0·48 (0·0153) (0·0066*)
3	0·43	0·44	0·44	0·35	0·36	0·36
4	0·34	0·35	0·36	0·26	0·28	0·29
5	0·28	0·28	0·30	0·21	0·22	0·23
6	0·23	0·23	0·25	0·17	0·18	0·19
			$N = 100$			
2	0·67 (0·0569) (0·0352*)	0·67 (0·0569) (0·0352*)	0·68 (0·0569) (0·0352*)	0·60 (0·0120) (0·0066*)	0·59 (0·0120) (0·0066*)	0·60 (0·0120) (0·0066*)
3	0·56	0·56	0·56	0·48	0·49	0·49
4	0·48	0·48	0·48	0·40	0·41	0·41
5	0·42	0·42	0·42	0·34	0·35	0·36
6	0·36	0·37	0·37	0·30	0·30	0·31

Approximations based on: (i) distribution of $s^2_{\max.}/s^2_{\min.}$; (ii) equation (8); (iii) equation (9); (iv) equation (13).

Exact values of the significance levels (i.e. $\Pr\{R \leqslant R_\alpha\}$) corresponding to a few of the approximate significance limits are shown in parentheses. These values often differ markedly from the nominal significance levels. This may be ascribed, in several cases, to the discontinuity of the distribution of R. To illustrate this, the next possible value of the tail probability above or below $\Pr\{R \leqslant R_\alpha\}$ is also given marked with an asterisk, according as to whether $\Pr\{R \leqslant R_\alpha\}$ is less than or greater than the nominal significance level α.

6. An alternative approximation to the distribution of R

Turning now to the approximation developed in §3 we see that

$$R = \frac{\min n_j}{\max n_j} \doteqdot \frac{\min x_j - \bar{x} + (N/k)^{\frac{1}{2}}}{\max x_j - \bar{x} + (N/k)^{\frac{1}{2}}}. \tag{10}$$

Taking logarithms, we have

$$\log_e R \doteqdot \log_e \left[1 + \frac{\min x_j - \bar{x}}{(N/k)^{\frac{1}{2}}}\right] - \log_e \left[1 + \frac{\max x_j - \bar{x}}{(N/k)^{\frac{1}{2}}}\right]. \tag{11}$$

The right-hand side of equation (11) may be expanded in powers of $N^{-\frac{1}{2}}$ to give

$$\log_e R \doteqdot -(k/N)^{\frac{1}{2}} (\max x_j - \min x_j) + O(N^{-1}). \tag{12}$$

Neglecting terms of order less than or equal to N^{-1} we obtain approximation (iv)

$$R_\alpha \doteqdot \exp[-w_\alpha (k/N)^{\frac{1}{2}}], \tag{13}$$

where w_α is the upper 100α % point of the distribution of the range of k independent unit normal variables.

Generally, each of the four approximations appears to give rather too large a value for R_α, approximation (iv) being the least accurate. However, it can be seen from the values of the probabilities given in Table 2 that there is appreciable discontinuity in the distribution of R, even when $N = 100$. This means that in several cases the exact values of the actual significance levels corresponding to the approximate significance limits are in fact the closest possible values to the nominal level α.

The difficulty of dealing with the discontinuities is shown by the figures in Table 3, where true probabilities are given for certain parts of the distribution when $N = 25$ and $k = 2, 3$. Arrows mark the position of the lower 5 and 1 % significance levels given by approximation (i).

Table 3. *Exact distributions of R for* $N = 25$, $k = 2, 3$

$k = 2$		$k = 3$			
R^*	$\Pr\{R \leqslant R^*\}$	R^*	$\Pr\{R \leqslant R^*\}$	R^*	$\Pr\{R \leqslant R^*\}$
0·136	0·0002	0·077	0·0012	0·187	0·0119
·190	·0009	·083	·0015	·200	·0151
·250	·0041	·111	·0016	·214	·0203
·316	·0146	·118	·0017	·231→	·0283
·389→	·0433	·125	·0022	·235	·0286
0·471→	·1078	·133	·0033	·250	·0409
·563	·2295	·143	·0049	·267	·0457
·667	·4244	·154	·0074	·273	·0516
·786	·6900	·167	·0103	·286	·0620
·924	1·0000	·176	·0107	·308→	·0802

Table 4. *Approximate upper significance limits for* $W = \max(n_j/N) - \min(n_j/N)$

(Actual significance levels in parentheses.)

	$N = 10$		$N = 25$	
k	$\alpha = 0{\cdot}05$	$\alpha = 0{\cdot}01$	$\alpha = 0{\cdot}05$	$\alpha = 0{\cdot}01$
2	0·619 (0·0215) (0·1094*)	0·814 (0·0020) (0·0215*)	0·0392 (0·0433) (0·1078*)	0·515 (0·0146) (0·0041*)
3	0·604 (0·0320) (0·0846*)	0·752 (0·0056) (0·0320*)	0·382 (0·0487) (0·0920*)	0·476 (0·0116) (0·0048*)
4	0·574 (0·0501) (0·0113*)	0·696 (0·0113) (0·0017*)	0·363 (0·0317) (0·0703*)	0·440 (0·0128) (0·0046*)
5	0·546 (0·0293) (0·0707*)	0·650 (0·0041) (0·0292*)	0·345	0·411
6	0·520	0·615	0·329	0·389

	$N = 50$		$N = 100$	
	$\alpha = 0{\cdot}05$	$\alpha = 0{\cdot}01$	$\alpha = 0{\cdot}05$	$\alpha = 0{\cdot}01$
2	0·277 (0·0649) (0·0328*)	0·364 (0·0066) (0·0153*)	0·196 (0·0569) (0·0352*)	0·257 (0·0120) (0·0066*)
3	0·270	0·336	0·191	0·238
4	0·257	0·311	0·182	0·220
5	0·244	0·291	0·173	0·206
6	0·233	0·275	0·165	0·194

7. Approximate upper significance limits of the distribution of range of relative frequencies

As a further application of the approximation developed in § 3, we will consider the distribution of

$$W = \max(n_j/N) - \min(n_j/N).$$

Using the results obtained in § 3, we see that

$$W \doteqdot w(kN)^{-\frac{1}{2}}, \tag{14}$$

where w is the range of k independent unit normal variables.

Table 22 of Pearson & Hartley (1954) gives significance limits for w. These have been used in (14) to give approximate values of the upper 5 and 1 % points of the distribution of W, as given in Table 4.

Exact values of the actual significance levels (i.e. $\Pr\{W \geqslant W_\alpha\}$) corresponding to some of the approximate significance limits are shown in parentheses. The second figure, marked with an asterisk, is the next possible value of the tail probability, above or below $\Pr\{W \geqslant W_\alpha\}$ according as to whether $\Pr\{W \geqslant W\alpha\}$ is less than or greater than the nominal level α.

The exact values of the actual significance levels, corresponding to the approximate significance limits, often differ markedly from the nominal level, α. However, the differences in the pairs of values of the tail probabilities, given in Table 4, indicate that this is almost entirely due to the discontinuity of the distribution of W. In fact, for every case in which the tail probabilities are given, the approximate significance limits in Table 4, would give practical procedures for which the difference, between the actual significance level ($\Pr\{W \geqslant W_\alpha\}$) and the nominal level α, is the minimum possible. Remembering this, the approximation appears to be reasonably adequate.

References

Bartlett, M. S. & Kendall, D. G. (1946). The statistical analysis of variance heterogeneity and the logarithmic transformation. *J.R. Statist. Soc.*, B, **8**, 128–38.

Eisenhart, C., Hastray, M. W. & Wallis, W. A. (1947). *Selected Techniques of Statistical Analysis* (Statistical Research Group, Columbia University). London and New York: McGraw Hill.

Johnson, N. L. (1960). An approximation to the multinomial distribution: Some properties and applications. *Biometrika*, **47**, 93–102.

Kozelka, R. M. (1956). Approximate upper percentage points for extreme values in multinomial sampling. *Ann. Math. Statist.* **27**, 507–12.

Pearson, E. S. & Hartley, H. O. (1954). *Biometrika Tables for Statisticians*, **1**. Cambridge University Press.

Young, D. H. (1960). Unpublished Ph.D. thesis. University of London.

Some distributions arising in the study of generalized mean differences

By T. A. RAMASUBBAN

London School of Economics and Political Science

1. In my first paper (1958) on mean differences, I defined the coefficient of mean differences of rth order as

$$\Delta_r = \int_{-\infty}^{\infty}\int |x-y|^r\, p(x)\, p(y)\, dx\, dy \tag{1·1}$$

(with integrals replaced by summations for discrete distributions). This is a generalized version of the usual Δ_1. I obtained the value of Δ_1 explicitly in the case of certain discontinuous populations such as positive and negative binomial, the Poisson, geometric and so on. Later (1959), I proceeded to consider the quantity Δ_r itself for the binomial and the Poisson distributions and, using certain functional properties found to be obeyed by (1·1) in these instances, evaluated in reasonably compact form, the expressions for Δ_r which are valid for all integral values of r. It was shown that whereas the even order mean differences can be determined very simply from the relation

$$\Delta_{2r} = 2\mu_{2r} + \sum_{j=2}^{2r-2} (-1)^j \binom{2r}{j} \mu_{2r-j}\mu_j, \tag{1·2}$$

where the μ's are the mean-moments of the distribution in question, the odd order Δ's are more difficult to express explicitly. For the binomial distribution, they are given by

$$\Delta_{2r+1} = 2\sum_{s=1}^{r+1} \left\{ a_{rs}\, n^{(s)} \binom{n-s}{s} p^{2s}q^{2s}\, {}_2F_1\{-n+2s, \tfrac{1}{2}(2s+1);\, 2s+1;\, 4pq\} \right.$$
$$\left. + a'_{rs}\, n^{(s)} \binom{n-s}{s-1} p^{2s-1}q^{2s-1}\, {}_2F_1\{-n+2s-1, \tfrac{1}{2}(2s-1);\, 2s-1;\, 4pq\} \right\}, \tag{1·3}$$

where $n^{(s)} = n(n-1)(n-2)\ldots(n-s+1)$, the hypergeometric function is given by the finite series

$$ {}_2F_1(-\alpha, \beta; \gamma; x) = 1 - \frac{\alpha\beta}{\gamma}\frac{x}{1!} + \frac{\alpha(\alpha-1)\beta(\beta+1)}{\gamma(\gamma+1)}\frac{x^2}{2!} - \ldots $$

and a_{rs} and a'_{rs} are constants whose numerical values have been tabulated for the first few r and s. For the Poisson distribution the corresponding expressions are

$$ \Delta_{2r+1} = 2e^{-2\lambda}\sum_{s+1}^{r+1}\lambda^s\{a_{rs}I_s(2\lambda) + a'_{rs}I_{s-1}(2\lambda)\}, \tag{1·4} $$

where $I_\nu(2\lambda)$ is the νth order Bessel function of the first kind with argument 2λ.

In this paper, I examine the general problem of expressing the Δ's in explicit form from a different point of view.

2. The underlying idea behind the present method follows from definition (1·1) for Δ_r. Defined as it is, it may be considered as the expectation of $|x-y|^r$, or in other words, as the rth absolute moment of a density function whose stochastic variable d is the algebraic difference of two variables x and y which follow independently some given parent distributions $p(x)$ and $p(y)$.

3. In order to enable a coherent examination of the Δ_r's derived already, we shall discuss here the application of this approach to the three important cases which we have considered earlier, viz. in which x and y have (i) the binomial, (ii) the Poisson and lastly (iii) the geometric distribution.

(i) The binomial parent

Since x and y are independent and similarly distributed in the binomial fashion $\binom{n}{t}p^tq^{n-t}$, the combined probability is

$$ p(x,y) = \binom{n}{x}\binom{n}{y}p^{x+y}q^{2n-x-y}. $$

The distribution $p_n(d)$ of the difference $d = (x-y)$, $d \geqslant 0$ is therefore obtained from

$$ p_n(d) = \sum_{y=0}^{n-d}\binom{n}{y}\binom{n}{y+d}p^{2y+d}q^{2n-2y-d}. \tag{3·1} $$

The sum in (3·1) is the coefficient of t^{-d} in $\{(q+pt)(q+p/t)\}^n$, i.e. in $\{1+\{(1-t)^2/t\}pq\}^n$ and this, when expressed in terms of the hypergeometric function ${}_2F_1$, leads to

$$ p_n(d) = \binom{n}{d}(pq)^d\,{}_2F_1\{-n+d, \tfrac{1}{2}(2d+1); 2d+1; 4pq\}, \tag{3·2} $$

If, however, d is negative and equal to $-d'(d' > 0)$, y and d in (3·1) are to be simply replaced by x and d', respectively, so that the summation now with respect to x leaves the final expression for the form of $p_n(d')$ exactly identical with (3·2). Thus the difference d, which takes all the integral values between $-n$ and $+n$, has a distribution symmetrical about zero. The odd moments of $p_n(d)$ are therefore zero and its even moments are given by the mean differences Δ_r of the binomial distribution.

Let us then consider this mean difference to be denoted by ${}_n\Delta_r$. We have

$$ {}_n\Delta_r = E|d|^r = 2\sum_{d=0}^{r} d^r p_n(d) \tag{3·3} $$

irrespective of whether r is odd or even. Substituting for $p_n(d)$ from (3·2) and using a result connecting the contiguous ${}_2F'_1$, namely,

$$ \begin{aligned} {}_2F_1\{-n+d, \tfrac{1}{2}(2d+1); 2d+1; 4pq\} &= {}_2F_1\{-n+d, \tfrac{1}{2}(2d-1); 2d-1; 4pq\} \\ &\quad - \frac{(n-d)(n-d-1)}{d(d+1)}(pq)^2\,{}_2F_1\{-n+d+2, \tfrac{1}{2}(2d+3); 2d+3; 4pq\}, \end{aligned} \tag{3·4} $$

which provides the main key to the summation in (3·3), it can be written, after a suitable rearrangement of the resulting terms, as

$$ \begin{aligned} {}_n\Delta_r &= 2npq\left[\sum_{d=-1}^{n-1}(d+1)^{r-1}p_{n-1}(d) - \sum_{d=1}^{n-1}(d-1)^{r-1}p_{n-1}(d)\right] \\ &= 2npq\left[(-1)^{r-1}p_{n-1}(0) + 2\sum_{j=0}^{[\frac{1}{2}(r-1)]}\binom{r-1}{2j+1}\sum_{d=0}^{n-1}d^{r-2j-2}p_{n-1}(d)\right] \\ &= 2npq\left[(-1)^{r-1}\,{}_2F_1(-n+1, \tfrac{1}{2}; 1; 4pq) + \sum_{j=0}^{[\frac{1}{2}(r-1)]}\binom{r-1}{2j+1}_{(n-1)}\Delta_{r-2j-2}\right], \end{aligned} \tag{3·5} $$

where $[\frac{1}{2}(r-2)]$ is, as usual taken to mean the greatest integer contained in the fraction $\frac{1}{2}(r-2)$.

This is the recurrence relation connecting Δ_r with $\Delta_{r-1}, \Delta_{r-2}, \ldots, \Delta_0$ and is valid for $r > 1$.

The case of $r = 1$ can be derived likewise with the help of (3·4) or be obtained from (1·3) by putting $r = 0$ therein, thus giving in either case

$$ {}_n\Delta_1 = 2npq[{}_2F_1(-n+1, \tfrac{1}{2}; 1; 4pq) + (n-1)\,pq\,{}_2F_1(-n+2, \tfrac{3}{2}; 3; 4pq)]. \tag{3·6} $$

Further, by distinguishing the even order differences from the rest, (3·5) can be shown to reduce to a simpler form. For, we then have,

$$ \begin{aligned} {}_n\Delta_{2r} &= 2npq\left[-{}_2F_1(-n+1, \tfrac{1}{2}; 1; 4pq) + \sum_{j=0}^{r-1} \binom{2r-1}{2j+1} {}_{(n-1)}\Delta_{2r-2j-2} \right] \\ &= 2npq\left[-{}_2F_1(-n+1, \tfrac{1}{2}; 1; 4pq) + (n-1)\,\Delta_0 + \sum_{j=0}^{r-2} \binom{2r-1}{2j+1} {}_{(n-1)}\Delta_{2r-2j-2} \right] \end{aligned} \tag{3·7} $$

and from the definition (3·3) for ${}_n\Delta_r$, ${}_n\Delta_0 = 2\sum_{d=0}^{n} p_n(d)$ which sums up to the value $1 + {}_2F_1(-n, \tfrac{1}{2}; 1; 4pq)$ on account of the fact that $p_n(d)$ is symmetric and that $\sum_{d=-n}^{n} p_n(d) = 1$. Hence, from (3·7)

$$ {}_n\Delta_{2r} = 2npq\left[1 + \sum_{j=0}^{r-2} \binom{2r-1}{2j+1} {}_{(n-1)}\Delta_{2r-2j-2} \right]. \tag{3·8} $$

Thus for $r = 1$, ${}_n\Delta_2 = 2npq$;

$r = 2$, ${}_n\Delta_4 = 2npq\{1 + 3\,{}_{(n-1)}\Delta_1\} = 2npq\{1 + 6(n-1)\,pq\}$

and so on. The same results, of course, can be obtained from (1·2) after substituting the appropriate moments of the binomial distribution.

In addition to (3·5) and (3·8), a specific relation for Δ_1 in terms of $(n-1)$, n and $(n+1)$ is also worth noting. For this purpose we imply the result

$$ \begin{aligned} &(4npq - 2pq - 2n)\,{}_2F_1(-n+1, \tfrac{1}{2}; 2; 4pq) + (n+1)\,{}_2F_1(-n, \tfrac{1}{2}; 2; 4pq) \\ &\qquad - (n-1)\,(4pq-1)\,{}_2F_1(-n+2, \tfrac{1}{2}; 2; 4pq) = 0 \end{aligned} \tag{3·9} $$

in combination with (3·6) which reduces to

$$ {}_n\Delta_1 = 2npq\,{}_2F_1(-n+1, \tfrac{1}{2}; 2; 4pq). \tag{3·10} $$

By substituting for ${}_2F_1$ in (3·9) the corresponding Δ_1's from (3·10) and simplifying we have the required result

$$ {}_{(n+1)}\Delta_1 + \{4pq - 2 - (2/n)\,pq\}\,{}_n\Delta_1 - (4pq-1)\,{}_{(n-1)}\Delta_1 = 0, $$

or

$$ {}_{(n+1)}\Delta_1 = (4pq-1)\,\{{}_{(n-1)}\Delta_1 - {}_n\Delta_1\} + \left(1 + \frac{2pq}{n}\right) {}_n\Delta_1. \tag{3·11} $$

In particular if $p = q = \frac{1}{2}$, (3·11) gives ${}_{(n+1)}\Delta_1 = \{1 + (1/2n)\}\,{}_n\Delta_1$. This relation was suggested without a complete proof by Gini (1912).

(ii) The Poisson parent

Once we have the distribution $p_n(d)$ for the binomial case, the distribution $p(d)$ for the Poisson can be readily derived by the usual method of considering the limit of the former as $n \to \infty$, $p \to 0$ and $np \to \lambda$.

Since under these limiting conditions

$$ {}_2F_1(-n+\alpha, \beta; \gamma; 4pq) \to {}_1F_1(\beta; \gamma; -4\lambda), $$

$$ p_n(d) = \binom{n}{d} (pq)^d\,{}_2F_1\{-n+d, \tfrac{1}{2}(2d+1); 2d+1; 4pq\}, $$

$$ p(d) = \frac{\lambda^d}{d!}\,{}_1F_1\{\tfrac{1}{2}(2d+1); 2d+1; -4\lambda\}, $$

or in terms of $I_\rho(2\lambda)$ (defined earlier), $p(d) = e^{-2\lambda} I_d(2\lambda)$. (3·12)

Irwin (1937) obtains (3·12) from the joint distribution of the two independent variates x and y, each following the Poisson law,

$$ p(t) = \frac{e^{-\lambda}\,\lambda^t}{t!} \quad (t = x, y). $$

Once again, (3·12) is symmetrical about zero since $I_{-d}(2\lambda) = I_{+d}(2\lambda)$ for all integral d.

The procedure of establishing the linear relationship between the Δ_r's is the same as before and is now accomplished by using the fact that

$$dI_d = \lambda(I_{d-1} - I_{d+1}), \quad \text{where} \quad I_d \equiv I_d(2\lambda). \tag{3·13}$$

Hence

$$\Delta_r = 2e^{-2\lambda} \sum_{d=0}^{\infty} d^r I_d$$

$$= 2\lambda e^{-2\lambda} \sum_{d=0}^{\infty} d^{r-1}(I_{d-1} - I_{d+1}),$$

which, after suitable adjustments, finally gives

$$\Delta_r = (-1)^{r-1} 2\lambda e^{-2\lambda} I_0 + 2\lambda \sum_{j=0}^{[\frac{1}{2}(r-2)]} \binom{r-1}{2j+1} \Delta_{r-2j-2} \tag{3·14}$$

for $r > 1$. For $r = 1$, the use of (3·13) leads to the known result

$$\Delta_1 = 2\lambda e^{-2\lambda}(I_0 + I_1) \tag{3·15}$$

(also obtained by putting $r = 0$ in (1·4)) so that Δ_r of any given order can be very conveniently built up with the help of (3·14) and (3·15).

In case of even r, however, a simpler relation exists, and we can show as in (3·8), that

$$\Delta_{2r} = 2\lambda\left[1 + \sum_{j=0}^{r-2} \binom{2r-1}{2j+1} \Delta_{2r-2j-2}\right]. \tag{3·16}$$

In particular,

$$\Delta_2 = 2\lambda, \quad \Delta_4 = 2\lambda(1 + 3\Delta_2) = 2\lambda(1 + 6\lambda)$$

and so on, which also constitute the even mean-moments of the difference distribution $p(d)$ and agree with Irwin's expressions.

(iii) The geometric parent

Finally, let us examine the case of geometric probability

$$p(t) = qp^t; \quad q + p = 1.$$

Starting with the joint probability $p(x, y) = q^2 p^{x+y}$ it can be shown as in (i) that the distribution of the difference $d = x - y$, follows the law,

$$p(d) = \frac{q^2}{1-p^2} p^{|d|} \quad (-\infty \leqslant d \leqslant \infty). \tag{3·17}$$

Thus the odd moments of $p(d)$ vanish and the even moments of order $2r$ are identical with Δ_{2r}.

The explicit values for the generalized mean differences Δ_r are given by

$$\Delta_r = \frac{2q^2}{1-p^2} \sum_{d=0}^{\infty} d^r p(d)$$

$$= \frac{2q^2 p}{1-p^2} \frac{1}{(1-p)^{r+1}} \sum_{s=1}^{r} A_{rs} p^{s-1} \tag{3·18}$$

$$= \frac{2q^2 p}{(1-p^2)\, q^{r+1}} A_r(p) \text{ say}, \tag{3·19}$$

where the A_{rs} are the Eulerian numbers. From (3·18), we thus have

$$\Delta_1 = \frac{2q^2 p}{1-p^2} \frac{1}{q^2} = \frac{2p}{1-p^2},$$

$$\Delta_2 = \frac{2q^2 p}{1-p^2} \frac{1+p}{q^3} = \frac{2p}{q^2},$$

$$\Delta_3 = \frac{2q^2 p}{1-p^2} \frac{1+4p+p^2}{q^4} = \frac{2p(1+4p+p^2)}{q^2(1-p^2)}, \text{ etc.}$$

The connexion between these Δ's can be established either directly from (3·18) or from the relationship

$$A_r(p) = rA_{r-1}(p) + (1-p)\frac{d}{dp}A_{r-1}(p). \tag{3·20}$$

On substituting the relevant expressions for A's from (3·19), (3·20) gives

$$\Delta_r = \left(\frac{2p}{1-p^2} + p\frac{d}{dp}\right)\Delta_{r-1}, \tag{3·21}$$

which unlike the previous relations is in the form of a difference-differential equation.

4. On the whole therefore, the indirect method described here seems worthwhile in enabling us (i) to obtain the Δ's in forms capable of easy and systematic handling and (ii) to study simultaneously, by identifying the moment-constants of $p(d)$ with Δ's, these secondary distributions, some of which as we have seen, belong to the special transcendental type. Generalizations and extensions of the above results may be attempted in many ways, as for example, by assuming that x and y have the same discrete forms but with different parameters or that they belong to two different discrete parents. Similar variations are possible for the continuous cases also.

References

Gini, C. (1912). *Variabilità e Mutabilità.*

Irwin, J. O. (1937). The frequency distribution of the difference between two independent variates following the same Poisson distribution. *J.R. Statist. Soc.* **100**, 415.

Ramasubban, T. A. (1958). The mean differences and the mean deviations of some discontinuous distributions. *Biometrika*, **45**, 549–56.

Ramasubban, T. A. (1959). The generalized mean differences of the binomial and Poisson distributions. *Biometrika*, **46**, 223–29.

On certain characteristics of some discrete distributions

By GEORGE E. BARDWELL

University of Denver

1. This paper stems from the following remark by Ramasubban (1958) on the mean deviation of some discontinuous distributions: 'Dr Johnson. . . has pointed out a remarkable general result which includes the hypergeometric, the binomial, the Poisson and the geometric distributions.' The result is

mean deviation (M.D.) = 2 (variance) (frequency function at $[\mu]$), where $[\mu]$ is the greatest integer not greater than the mean. (1)

The statement is, in fact, true for the binomial and the Poisson, true for the geometric distribution for integral values of μ but does not hold for the hypergeometric.* In considering it I was led to some properties of discontinuous frequency functions which are presented below.

2. Let $f(x, n, p)$ be a discrete probability density function of x ($x = 0, 1, 2, \ldots$), with parameters n and p where p may be continuous in the interval $0 \leqslant n \leqslant a \leqslant \infty$. Let μ be the mean of the distribution and μ_i the ith moment about the mean. μ is then a function of n and p, say $g(n, p)$. Suppose that for fixed n, $g(n, p)$ is continuous and strictly increasing in the interval $(0, a)$. Then it has a continuous strictly increasing inverse $h(n, \mu) = g^{-1}(n, p) = p$. For each x let $\partial f/\partial p$ and $\partial h/\partial p$ exist in $(0, a)$.

* *Editorial Note.* Mr Ramasubban points out that for the hypergeometric

$$f(x) = \binom{Np}{x}\binom{Nq}{n-x}\Big/\binom{N}{n}$$

Johnson's relation holds except for a term N instead of $N-1$ in the denominator, i.e.

$$\delta_1 = f(m)\frac{2m(Nq-n+m)}{N},$$

where $m = [\mu]$.

THEOREM 1. If f obeys the relation

$$\frac{\partial f}{\partial \mu} = \frac{\partial f}{\partial p}\frac{dp}{d\mu} = w(n,p)\{x - v(n,p)\}f, \tag{1}$$

then (a)
$$v(n,p) = \mu, \tag{2}$$

(b)
$$w(n,p) = 1/\mu_2, \tag{3}$$

(c)
$$\mu_i = \mu_2\left\{\frac{d\mu_{i-1}}{d\mu} + (i-1)\mu_{i-2}\right\}. \tag{4}$$

We have
$$\sum_{x=0}^{\infty}\frac{\partial f}{\partial \mu} = \frac{d}{d\mu}\Sigma f = 0. \tag{5}$$

Hence, from (1)
$$\Sigma w(n,p)\{x - v(n,p)\}f = 0$$

giving (a) immediately. Since
$$\frac{\partial}{\partial \mu}(x-\mu)f = (x-\mu)\frac{\partial f}{\partial \mu} - f$$

we have, using (1) with
$$\mu = v(n,p),$$
$$\frac{\partial}{\partial \mu}(x-\mu)f = w(n,p)(x-\mu)^2 f - f.$$

Summing over x we find
$$0 = w(n,p)\mu_2 - 1,$$

where (b) follows.

To establish (c), write

$$\frac{\partial}{\partial \mu}(x-\mu)^{i-1}f = -(i-1)(x-\mu)^{i-2}f + (x-u)^{i-1}\frac{\partial}{\partial \mu}f$$
$$= -(i-1)(x-u)^{i-2}f + \frac{1}{\mu_2}(x-\mu)^i f.$$

Summation of both sides for x yields (c).

3. A characteristic of the family of functions obeying (1) is that each has a single maximum which occurs when $x = \mu$. This property is shared by the binomial, the negative binomial, the Poisson, the logarithmic and the geometric distributions. The iterative process (4) for finding the moments is well known for the binomial and the Poisson.

THEOREM 2. The density function satisfying equation (1) is of the form

$$f(x,n,p) = C(n,x)\exp\left\{\int \frac{x-\mu}{\mu_2(dp/d\mu)}\,dp\right\}. \tag{6}$$

In fact equation (1) may be written
$$\frac{\partial f}{\partial \mu} = \frac{x-\mu}{\mu_2}f.$$

For fixed n, $(x-\mu)/\{\mu_2(dp/d\mu)\}$ is a function of p alone and thus we have

$$\frac{\partial f}{\partial p} = \frac{x-\mu}{\mu_2(dp/d\mu)}f$$

as a linear equation of first order in p. The solution is that given in (6), $C(n,x)$ being a normalizing constant such that the total frequency is unity. It is readily verified that the distributions mentioned earlier in this section satisfy the relationship.

4. THEOREM 3.

For the family of distributions of equation (6),

$$\text{M.D.} = -2\mu_2\frac{\partial}{\partial \mu}F([\mu],n,p), \tag{7}$$

where
$$F([\mu],n,p) = \sum_{x=0}^{[\mu]} f(x,n,p). \tag{8}$$

We have
$$\sum_{x=0}^{\infty}(x-\mu)f = -\sum_{x=0}^{[\mu]}(\mu-x)f + \sum_{x=[\mu]+1}^{\infty}(x-\mu)f = 0.$$

Then
$$\text{M.D.} = \sum_{0}^{\infty}|x-\mu|f = 2\sum_{0}^{[\mu]}(\mu-x)f. \tag{9}$$

Substituting from Theorem (1) the expression

$$\mu_2\frac{\partial f}{\partial \mu} = (x-\mu)f,$$

we have
$$\text{M.D.} = -2\mu_2\sum_{x=0}^{[\mu]}\frac{\partial f}{\partial \mu}$$
$$= -2\mu_2\frac{\partial}{\partial \mu}F([\mu], n, p)$$

as given at (7).

5. A further result can be obtained for certain distributions, such as the Poisson, which obey an additional restriction.

THEOREM 4.

If the members of the family of distributions obeying equation (1) also obey

$$\frac{d}{d\mu}f(x,p) = f(x-1,p) - f(x,p) \quad (x = 1, 2, \ldots). \tag{10}$$

then (*a*)
$$\text{M.D.} = 2\mu_2 f([\mu], p) + 2(\mu-\mu_2)f(0,p), \tag{11}$$

(*b*)
$$f(x,p) = f(0,p)\,\frac{\mu_2^x\Gamma(1+\mu_2-\mu)}{\Gamma(x+1+\mu_2-\mu)}. \tag{12}$$

with
$$f(0,p) = \frac{1}{M(1,\ 1+\mu_2-\mu,\ \mu_2)}$$

where M is the confluent hypergeometric function.

We have directly
$$\text{M.D.} = 2\sum_{x=0}^{[\mu]}(\mu-x)f$$
$$= 2\sum_{x=1}^{[\mu]}(\mu-x)f + 2\mu f(0,p).$$

Using Theorem 1 and (10) we then have

$$\text{M.D.} = 2\sum_{x=1}^{[\mu]}\mu_2\{f(x,p) - f(x-1,p)\} + 2\mu f(0,p)$$
$$= 2\mu_2 f([\mu], p) + 2(\mu-\mu_2)f(0,p)$$

as given at (11). Further, replacing x by $x+1$ in (10) we have

$$f(x,p) - f(x+1,p) = \frac{x+1-\mu}{\mu_2}f(x+1,p). \tag{13}$$

Regarding this as a difference equation we readily find equation (12).

REFERENCE

RAMASUBBAN, T. A. (1958). The mean difference and the mean deviation of some discontinuous distributions. *Biometrika*, **45**, 549–56.

Approximate tests for m rankings

By H. LINHART*

National Institute for Personnel Research, South African Council for Scientific and Industrial Research

1. Introduction

One or two sets of rankings of n objects are given. The sets, of size m_1 and m_2, are regarded as random samples from a population of rankings. Approximate tests for the following hypotheses are described below.

For one set of rankings

H. 1: the mean ranks of a certain subset of the n objects are identical.

For two sets of rankings

H. 2: the n mean ranks (generally different within a population) are identical in both populations.

H. 3: the coefficients of concordance are identical in both populations.

The investigation which gave rise to the development of the tests might be sketched here for an illustration. Bantu pupils have been asked to rank 6 different jobs in a number of job families according to their 'preferences'. The object of the study was to get an impression of the occupational ambitions of Bantu youths, to find out whether there is a common job hierarchy, and if so whether it changes with educational standard of the youths.

Coefficients of concordance were computed within the three educational groups (Standard VI, Form III, Form V) and in most cases found to be significant. But a more thorough statistical analysis is not possible without tests of differences between educational groups (H. 2) and tests whether a hierarchy obtained for some group is really significant in all its parts (H. 1). Information whether agreement of opinion within all educational groups is equal or not would be of additional value (H. 3).

The following terminology will be used: *a ranking* is a row containing the n *ranks*, $1, 2, \ldots, n$, in some order. A fixed *distribution of rankings*, specified by the probabilities for the $n!$ different rankings, simultaneously determines the n-variate *distribution of ranks*.

The methods employed are those used by Stuart (1951). No assumptions are made about the distribution of rankings. Suitable statistics are chosen and the first moments of the statistics are calculated as functions of the cumulants of the distribution of ranks. The expressions for the moments are simplified by neglecting cumulants above a certain order, and the cumulants left in those expressions are replaced by their k-statistics. The approximate moments so obtained are used to fit some simple distribution for the statistic under the null hypothesis, and the test is carried out.

Stuart (1951) obtained four moments and then used Karl Pearson's criterion to decide on the most appropriate type of distribution to fit. This would be too complex a procedure to carry out here. We calculated therefore only the first two moments and suggest fitting the simplest tabulated distributions: the normal distribution for statistics ranging from $-\infty$ to $+\infty$, and the χ^2-distribution for statistics ranging from 9 to ∞.

The procedure described above yields, of course, only very rough results and has to be applied with sense. Stuart (1951) developed and applied successfully similar tests which involve the same degree of approximation. This and the fact that no better methods are available seems to justify putting forward further crude methods.

The crux of the matter is that too many parameters enter the problem ($n!$ minus one) if *any* distribution of rankings is admitted. As Kendall remarked (1950), the major outstanding problem of ranking theory is to find some way of specifying a population with a tractable number of parameters. As long as that problem has not been solved one has to be satisfied with crude methods, similar to those suggested here.

2. Notation

The notation follows that of Stuart (1951). Cumulants and k-statistics are written in the customary notation (Fisher, 1928), but where possible zero indices are omitted. If not otherwise indicated all cumulants and k-statistics appearing in formulae or in computations are those of the distribution of

* Now with the African Wool Textile Research Institute, Rhodes University, Grahamstown.

ranks. Where necessary κ and k are used for cumulants and k-statistics in the one population and λ and l for cumulants and k-statistics in the other population. Summation is always over all relevant variables ('objects'), *permutations are not excluded*. To give an example: for three objects (a three-variate distribution of ranks) $\Sigma\Sigma(k_{10}-k_{01})^2$ is identical with $2[(k_{100}-k_{010})^2+(k_{010}-k_{001})^2+(k_{100}-k_{001})^2]$. The notation is easily understood if one reads $\Sigma\Sigma(k_{10}-k_{01})^2$ as $\sum_{i,j;\,i\neq j}^{n}(k_{10}^{(ij)}-k_{01}^{(ij)})^2$, where $k_{\alpha\beta}^{(ij)}$ is written for a k-statistic with n indices, the ith index being α, the jth index being β, all other indices being zero.

3. An extension of Stuart's test

Stuart (1951) gave a test for differences in mean ranks between any two of the n objects which have been ranked. His test statistic is $k_{10}-k_{01}$, the observed difference in mean ranks of the 2 objects. If the equality of more than two mean ranks has to be tested, a reasonable statistic seems to be

$$t=\Sigma\Sigma(k_{10}-k_{01})^2, \tag{1}$$

where the summation extends over all objects which have to be tested.

The mean and variance of t may be obtained in the following way. $E(t)$ and $E(t^2)-E^2(t)$ are expressed in terms of moments of the multivariate distribution of the *mean ranks* k_1. The moments are then expressed in terms of cumulants, and finally the cumulants of the distribution of *mean ranks* are calculated as functions of cumulants of the distribution of *ranks*. Here, where only first order k-statistics are involved, the last step is easily done because

$$\kappa_{rstu\ldots}(k_1)=\frac{\kappa_{rstu\ldots}}{m^{r+s+t+u+\ldots-1}}, \tag{2}$$

where $r, s, t, u, \ldots$ are zero or positive integers and where the cumulants on the left side are those of the distribution of mean ranks and on the right side are those of the distribution of ranks.

The actual calculations are very lengthy because of the complicated formulae connecting moments and cumulants. The methods and results given in Kendall (1940) and Cook (1951) were used whenever possible. The calculations are not displayed here and one may refer to Stuart (1951) who worked out in detail a few similar calculations.

The following results for mean and variance were obtained

$$E(t)=\frac{1}{m}\Sigma\Sigma(\kappa_{20}+\kappa_{02}-2\kappa_{11}), \tag{3}$$

$$\begin{aligned}V(t)\approx\frac{2}{m^2}\{&2\Sigma\Sigma(\kappa_{20}+\kappa_{02}-2\kappa_{11})^2\\&+4\Sigma\Sigma\Sigma(\kappa_{200}-\kappa_{110}-\kappa_{101}+\kappa_{011})^2\\&+\Sigma\Sigma\Sigma\Sigma(\kappa_{1010}-\kappa_{1001}-\kappa_{0110}+\kappa_{0101})^2\}.\end{aligned} \tag{4}$$

The exact variance contains cumulants up to order four, the formula given above is an approximation, where cumulants of higher order than two have been neglected, (this is equivalent to neglecting terms having a higher power than two of m in the denominator).

If the test has to be applied, the cumulants must be replaced by their k-statistics and a distribution must be fitted for t. It is proposed to use a χ^2-distribution with ν degrees of freedom for t/α. If this is done, one has to put

$$\alpha=\frac{V(t)}{2E(t)}, \tag{5}$$

$$\nu=\frac{2E^2(t)}{V(t)}=\frac{E(t)}{\alpha}, \tag{6}$$

in order to assure for t a mean and a variance as given in (3) and (4) above.

4. Test for equality of mean ranks between two groups

The proposed statistic is

$$t=\Sigma(k_1-l_1)^2. \tag{7}$$

The mean of t is

$$E(t)=\Sigma\left(\frac{\kappa_2}{m_1}+\frac{\lambda_2}{m_2}\right), \tag{8}$$

and the variance (exact)
$$V(t) = \Sigma\left[\frac{\kappa_4}{m_1^3}+\frac{\lambda_4}{m_2^3}+2\left(\frac{\kappa_2}{m_1}+\frac{\lambda_2}{m_2}\right)^2\right] + \Sigma\Sigma\left[\frac{\kappa_{22}}{m_1^3}+\frac{\lambda_{22}}{m_2^3}+2\left(\frac{\kappa_{11}}{m_1}+\frac{\lambda_{11}}{m_2}\right)^2\right]. \tag{9}$$

The approximate variance is
$$V(t) \approx 2\Sigma\left(\frac{\kappa_2}{m_1}+\frac{\lambda_2}{m_2}\right)^2 + 2\Sigma\Sigma\left(\frac{\kappa_{11}}{m_1}+\frac{\lambda_{11}}{m_2}\right)^2. \tag{10}$$

Here again a χ^2-distribution with ν degrees of freedom should be fitted for t/α, where ν and α are given by (6) and (5).

5. Test for equality of coefficients of concordance between two groups

We define w, the coefficient of concordance in the population, in analogy with W, the coefficient of concordance in the sample, as
$$w = \frac{12}{n(n^2-1)}\,\Sigma(\kappa_1 - \tfrac{1}{2}(n+1))^2. \tag{11}$$

Testing for equality of coefficients of concordance reduces thus to testing for equality of $\Sigma\kappa_1^2$.

The statistic used is
$$t = \Sigma(k_1^2 - l_1^2). \tag{12}$$

It has mean
$$E(t) = \Sigma\left(\frac{\kappa_2}{m_1}-\frac{\lambda_2}{m_2}\right), \tag{13}$$

variance (exact)
$$V(t) = \Sigma\left(\frac{\kappa_4}{m_1^3}+\frac{4\kappa_3\kappa_1+2\kappa_2^2}{m_1^2}+\frac{4\kappa_2\kappa_1^2}{m_1}+\frac{\lambda_4}{m_2^3}+\frac{4\lambda_3\lambda_1+2\lambda_2^2}{m_2^2}+\frac{4\lambda_2\lambda_1^2}{m_2}\right) + \Sigma\Sigma\left(\frac{\kappa_{22}}{m_1^3}+\frac{2\kappa_{10}\kappa_{12}+2\kappa_{01}\kappa_{21}+2\kappa_{11}^2}{m_1^2}+\frac{4\kappa_{11}\kappa_{01}\kappa_{10}}{m_1}+\frac{\lambda_{22}}{m_2^3}+\frac{2\lambda_{10}\lambda_{12}+2\lambda_{01}\lambda_{21}+2\lambda_{11}^2}{m_2^2}+\frac{4\lambda_{11}\lambda_{01}\lambda_{10}}{m_2}\right), \tag{14}$$

and approximate variance
$$V(t) \sim 4\Sigma\left(\frac{\kappa_2\kappa_1^2}{m_1}+\frac{\lambda_2\lambda_1^2}{m_2}\right)+4\Sigma\Sigma\left(\frac{\kappa_{11}\kappa_{01}\kappa_{10}}{m_1}+\frac{\lambda_{11}\lambda_{01}\lambda_{10}}{m_2}\right). \tag{15}$$

It is suggested that a normal distribution be used for t.

6. Examples of application

The k-statistics were computed using the following formulae
$$\left.\begin{aligned} k_1 &= \frac{s_1}{m}, \\ k_2 &= \frac{1}{m-1}\left(s_2-\frac{s_1^2}{m}\right), \\ k_{11} &= \frac{1}{m-1}\left(s_{11}-\frac{s_{10}s_{01}}{m}\right), \end{aligned}\right\} \tag{16}$$

where $s_{ij} = \sum^m x^i y^j$, and x and y are the ranks of the objects involved.

The following checks were used
$$\left.\begin{aligned} \Sigma s_1 &= \tfrac{1}{2}mn(n+1), \\ \Sigma s_2 &= \tfrac{1}{6}mn(n+1)(2n+1), \\ \Sigma\Sigma s_{11} &= (\Sigma s_1)^2/m - \Sigma s_2. \end{aligned}\right\} \tag{17}$$

The sum of the elements in the jth column of the complete s_{11} matrix (zeros in the diagonal): if m_i individuals give the jth object the rank i, the column sum is equal to $\sum_{i=1}^{n} m_i i\{\frac{1}{2}n(n+1)-i\}$. For higher moments similar checks may be worked out using, for example, David's & Johnson's formulae (1954, p. 239).

The test introduced in §3 above has been applied to Stuart's (1951) data. The following results were obtained in a *test for equality of mean ranks* of jobs *B*, *D* and *G*:

$$t = 8{\cdot}51, \quad E(t) = 0{\cdot}318,$$
$$V(t) = 0{\cdot}112, \quad \alpha = 0{\cdot}177,$$
$$\nu = 1{\cdot}80, \quad \chi^2 = t/\alpha = 48{\cdot}1.$$

For two degrees of freedom this value of χ^2 is highly significant, a result which agrees with Stuart's findings.

Among other results, the following observations were made in the investigation mentioned in the introduction.

Table of frequencies of ranks

		Form V						Standard VI					
	Ranks	1	2	3	4	5	6	1	2	3	4	5	6
A	Tradesman	1	4	5	6	11	3	3	7	6	6	6	2
B	Businessman	0	11	5	6	6	2	0	3	2	7	12	6
C	Professional man	28	1	1	0	0	0	13	5	5	2	2	3
D	Clerk	1	9	9	7	3	1	8	6	7	7	1	1
E	Public servant	0	5	9	8	6	2	6	6	10	6	1	1
F	Factory operative	0	0	1	3	4	22	0	3	0	2	8	17

Table of k-statistics

Form V

	k_1	k_2	A	B	C	D	E	F
A	4·033	1·829	—	0·1242	−0·1755	−0·8322	−0·9194	−0·0176
B	3·433	1·843	—	—	−·1134	−·6943	−·7608	−·3900
C	1·100	0·162	—	—	—	−·0172	·1000	·0448
D	3·167	1·383	—	—	—	—	·3263	−·1686
E	3·700	1·389	—	—	—	—	—	−·1345
F	5·567	0·664	—	—	—	—	—	—

Standard VI

	k_1	k_2	A	B	C	D	E	F
A	3·367	2·168	—	−0·0659	−1·2813	−0·5986	−0·3262	0·0948
B	4·533	1·433	—	—	−·8084	−·0566	−·0772	−·4189
C	2·467	2·945	—	—	—	−·0124	−·3711	−·4777
D	2·667	1·883	—	—	—	—	−·6676	−·5535
E	2·767	1·631	—	—	—	—	—	−·1949
F	5·200	1·544	—	—	—	—	—	—

Test for equality of mean ranks:

$$m_1 = m_2 = 30, \quad t = 4{\cdot}777,$$
$$E(t) = 0{\cdot}629, \quad V(t) = 0{\cdot}177,$$
$$\alpha = 0{\cdot}141, \quad \nu = 4{\cdot}46,$$
$$\chi^2 = t/\alpha = 33{\cdot}9,$$

a value which is highly significant for five degrees of freedom.

Test for equality of coefficients of concordance:

$$m_1 = m_2 = 30, \quad t = 4{\cdot}192,$$
$$E(t) = -0{\cdot}145, \quad SD(t) = 1{\cdot}553,$$
$$(t - E(t))/SD(t) = 2{\cdot}81,$$

which is significant at the 1 % level.

To get an impression of the degree of approximation in (15), the terms of (14) of order m^{-2} were added to (15) to give an improved approximation for the variance. After considerable computation the improved value for the standard deviation was found to be 1·532, which is only slightly different from 1·553 as obtained using formula (15).

It appears that the Form V and Standard VI pupils do not have the same job hierarchy and Form V pupils agree among themselves more closely than Standard VI pupils.

References

Cook, M. B. (1951). Bivariate k-statistics and cumulants of their joint sampling distribution. *Biometrika*, **38**, 179–95.

David, F. N. & Johnson, N. L. (1954). Statistical treatment of censored data. Part I. Fundamental formulae. *Biometrika*, **41**, 228–40.

Fisher, R. A. (1928). Moments and product moments of sampling distributions. *Proc. Lond. Math. Soc.* **30**, 199–238.

Kendall, M. G. (1940). The derivation of multivariate sampling formulae from univariate formulae by symbolic operation. *Ann. Eugen., Lond.*, **10**, 392–402.

Kendall, M. G. (1950). Discussion on symposium on ranking methods. *J. Roy. Statist. Soc.* B, **12**, 189.

Lake, J. R. & Mbau, G. G. (1960). Occupational preferences among male Bantu school children. To be published in *J. Nat. Inst. Personnel Res.*

Stuart, A. (1951). An application of the distribution of the ranking concordance coefficient. *Biometrika*, **38**, 33–42.

A note on the derivation of some exact multivariate tests

By A. M. KSHIRSAGAR

University of Manchester

Williams (1952) has derived the conditional distribution of the squares of the sample canonical correlations r_i^2 and the squares of the residual canonical correlations ϕ_j when the true discriminant function corresponding to the only non-zero canonical correlation, is eliminated. He has, however, not explicitly stated some of the arguments in the derivation. This note is intended to clarify these points, and treat the problem more generally.

We consider a dependent vector variate

$$\mathbf{x}' = (x_1, \ldots, x_p)$$

and an independent vector variate $\mathbf{y}' = (y_1, \ldots, y_q)$,

where $p \leqslant q$. These are taken to be the population canonical variables. The sample x_{it}, y_{jt} $(i = 1, p; \ldots, j = 1, \ldots, q; t = 1, \ldots, n)$ with n degrees of freedom is divided in the usual way (see Bartlett, 1934) into a part with q degrees of freedom corresponding to the regression of $\mathbf{x}$ on $\mathbf{y}$ and the remaining part with $n-q$ degrees of freedom. Let A and B be the matrices of the sum of squares and products corresponding to this division. Then they have independent Wishart distributions. This is so if the true canonical correlations ρ_i are all zero. If, however, ρ_1 is the only non-zero correlation, then A has a non-central Wishart distribution belonging to the 'linear' case (see Anderson, 1946). In any case, make the transformation (see Bartlett, 1947)

$$KAK' = WDW',$$

$$K(A+B)K' = WW',$$

where K is the diagonal matrix with diagonal elements, $(1-\rho_1^2)^{-\frac{1}{2}}, 1, \ldots, 1$; D is the diagonal matrix of the quantities r_i^2 and $W = [w_{ij}]$. Then the joint distribution of W and D (see Bartlett, 1947, equations (4) to (7)) when all the ρ's are zero is

$$C_1 |WW'|^{\frac{1}{2}(n-p)} \exp\left[-\frac{1}{2}\sum_{i,j=1}^{p} w_{ij}^2\right] dW \times C_2 \prod_{i=1}^{p}\left\{(r_i^2)^{\frac{1}{2}(q-p-1)}(1-r_i^2)^{\frac{1}{2}(n-q-p-1)} \prod_{j=i+1}^{p}(r_i^2 - r_j^2)\right\} \prod_{i=1}^{p} dr_i^2,$$

where

$$C_1 = 2^{p-\frac{1}{2}np}\pi^{-\frac{1}{4}p^2} \prod_{i=0}^{p-1} \{\Gamma[\tfrac{1}{2}(p-i)]/\Gamma[\tfrac{1}{2}(n-i)]\}$$

(the w_{ij} varying from $-\infty$ to ∞ except that $w_{1i} \geqslant 0$), and

$$C_2 = \pi^{\frac{1}{2}p} \prod_{i=0}^{p-1} \{\Gamma[\tfrac{1}{2}(n-i)]/(\Gamma[\tfrac{1}{2}(p-i)]\,\Gamma[\tfrac{1}{2}(q-i)]\,\Gamma[\tfrac{1}{2}(n-q-i)])\}.$$

If ρ_1 is the only non-zero true canonical correlation, there is an extra factor

$$e^{-\frac{1}{2}\lambda^2} \sum_{\alpha=0}^{\infty} \frac{(\frac{1}{2}\lambda^2)^\alpha}{\alpha!} \frac{(w_{11}^2 r_1^2 + \ldots + w_{1p}^2 r_p^2)^\alpha\, \Gamma(\frac{1}{2}q)}{2^\alpha\, \Gamma(\frac{1}{2}q+\alpha)}, \tag{1}$$

where

$$\lambda^2 = \rho_1^2 \sum_{t=1}^{n} y_{1t}^2/(1-\rho_1^2)$$

in the joint distribution of W and D. If $\mathbf{u}' = (u_1, \ldots, u_p)$ and $\mathbf{v}' = (v_1, \ldots, v_q)$ are the sample canonical variables corresponding to $\mathbf{x}'$ and $\mathbf{y}'$, then a little consideration will show that

$$\mathbf{x} = K^{-1}W\mathbf{u}$$

and hence $x_1 = (1-\rho_1^2)^{\frac{1}{2}}(w_{11}u_1 + \ldots + w_{1p}u_p)$ is the true discriminant function, in terms of the sample canonical variables. Williams (1952) has derived the distribution of ϕ_j, the squares of the residual canonical correlations, when the relation between x_1, the true discriminant function and $\mathbf{y}$ is eliminated. The relationship is measured by η, the square of the sample multiple correlation coefficient between x_1 and $\mathbf{y}$. This is obviously $(w_{11}^2 r_1^2 + \ldots + w_{1p}^2 r_p^2)/(w_{11}^2 + \ldots + w_{1p}^2)$. Williams has derived the distribution by transforming from w_{1i} $(i = 1, \ldots, p)$ to ϕ_j $(j = 1, \ldots, p-1)$ and $s = \sum_{i=1}^{p} w_{1i}^2$ in the distribution of the w_{1i}, which he simply takes as const. $\exp -\left(\frac{1}{2}\sum_{1}^{p} w_{1i}^2\right)$. But rigorously, as we have seen above, the distribution of w_{1i} must be obtained from the distribution of W, viz.

$$\text{const.}\,|WW'|^{\frac{1}{2}(n-p)}\, e^{-\frac{1}{2}\mathrm{tr}WW'}$$

multiplied by the extra factor given in (1) above, by integrating out the other w's. Thus, as Bartlett (1947) has observed, the distribution is normal, except for the 'linkage' factor, const. $|WW'|^{\frac{1}{2}(n-p)}$. For integrating out the other w's a transformation may be made to c_{ij} and θ_{ij} as in Bartlett's (1947) paper. It will then be observed, that the distribution of s is not a χ^2 with p degrees of freedom as given in Williams's (1952) paper, but is elevated to n degrees of freedom due to the 'linkage factor'. This has been observed by Bartlett (1947). However, since finally, for the distribution of ϕ's, s is to be integrated out, the desired distribution of r_i^2 and ϕ_j is as given in Williams's (1952) paper. It is

$$\frac{\text{const.}\,\Pi\{(r_i^2)^{\frac{1}{2}(q-p-1)}(1-r_i^2)^{\frac{1}{2}(n-q-2)}\}\,\big|\prod_{i\neq j}(r_i^2-r_j^2)\prod_{h\neq k}(\phi_h-\phi_k)\big|}{\Pi(1-\phi_k)^{\frac{1}{2}p}\,|\Pi(r_i^2-\phi_k)|^{\frac{1}{2}}} \times e^{-\frac{1}{2}\lambda^2} F(\tfrac{1}{2}n, \tfrac{1}{2}q, \tfrac{1}{2}\lambda^2\eta)\,\Pi dr_i^2\,\Pi d\phi_k,$$

where

$$(1-\eta)\,\Pi(1-\phi_k) = \Pi(1-r_i^2).$$

The distribution of η is $\quad \text{const.}\,\eta^{\frac{1}{2}(q-2)}(1-\eta)^{\frac{1}{2}(n-q-2)}e^{-\frac{1}{2}\lambda^2}F(\tfrac{1}{2}n, \tfrac{1}{2}q, \tfrac{1}{2}\lambda^2\eta)\,d\eta.$

The conditional distribution of r_i^2 and ϕ_k, when η is fixed, is therefore

$$\frac{\text{const.}\,\Pi(r_i^2)^{\frac{1}{2}(q-p-1)}\,\Pi(1-\phi_k)^{\frac{1}{2}(n-p-q-2)}\,\big|\prod_{i\neq j}(r_i^2-r_j^2)\prod_{h\neq k}(\phi_h-\phi_k)\big|\,\Pi dr_i^2\,\Pi d\phi_k}{\eta^{\frac{1}{2}(q-2)}\,|\Pi(r_i^2-\phi_k)|^{\frac{1}{2}}\,d\eta}$$

(One of the set r_i^2 and ϕ_k must be replaced by η, using the relation between them; but to preserve symmetry, this is not done.) This conditional distribution does not involve ρ_1; η is thus a sufficient statistic. This is why Williams (1952) is able to ignore the extra factor (1) and treat the w_{1i} as independent normal deviates.

Williams has derived the relation between ϕ and r_i^2 and w's by employing a geometrical argument involving principal semi-axes of an ellipsoid. This can alternatively be obtained, analytically, by observing that the ϕ's satisfy the determinantal equation

$$|-(1-\phi)\,C_{yy.x_1} + C_{yy.x}| = 0,$$

where the notation of Bartlett (1939) is used. When the sample canonical variables $\mathbf{u}$ and $\mathbf{v}$ are used, this becomes

$$|-(1-\phi)\,C_{vv.\Gamma u_1} + C_{vv.u}| = 0$$

as x_1, the true discriminant function, is proportional to $\mathbf{l}'\mathbf{u}$, where

$$\mathbf{l}' = (l_1, \ldots, l_p) \quad \text{and} \quad l_i = w_{1i}/(w_{11}^2 + \ldots + w_{1p}^2)^{\frac{1}{2}}.$$

Obviously, $C_{vv.u} = I - D$ and $C_{vv.l'u} = (\delta_{ij} - r_i r_j l_i l_j)$ where δ_{ij} is the Kronecker delta. Then the determinantal equation simplifies to

$$\sum_{i=1}^{p} \frac{w_{1i}^2(1-r_i^2)}{r_i^2 - \phi} = 0.$$

This is the required relation.

The author is indebted to Prof. M. S. Bartlett for much helpful advice.

References

Anderson, T. W. (1946). The non-central Wishart distribution and certain problems of multivariate analysis. *Ann. Math. Statist.* **17**, 409.

Bartlett, M. S. (1934). The vector representation of a sample. *Proc. Camb. Phil. Soc.* **30**, 327.

Bartlett, M. S. (1939). A note on tests of significance in multivariate analysis. *Proc. Camb. Phil. Soc.* **35**, 180.

Bartlett, M. S. (1947). The general canonical correlation distribution. *Ann. Math. Statist.* **18**, 1.

Williams, E. J. (1952). Some exact tests in multivariate analysis. *Biometrika*, **39**, 17.

Estimation from a linear Markov process

By N. Gilbert

John Innes Horticultural Institution, Bayfordbury, Hertford, Herts

This note is concerned with the feasibility of analysing growth curves as Markov chains. The idea was to transform to approximate linearity (and equality of error variance) and then fit a linear Markov process including a linear regression

$$y_i = m + b\{i - \tfrac{1}{2}(n+1)\} + x_i$$

where

$$x_i = r x_{i-1} + \epsilon_i,$$

$$\epsilon_1 \text{ is } N(0, \sigma^2), \quad \epsilon_i \text{ is } N(0, \overline{1-r^2}\sigma^2)$$

and n is the number of observations in the chain. This note describes the results of applying least-squares analysis (including the end-correction of Quenouille (1957) in the denominator of $\hat{r}$) to a set of artificial data. The least-squares equations were solved iteratively, continuing until $\hat{r}$ differed from its value in the previous cycle by less than 2^{-9}. Convergence is rapid except in a few cases for large values of r; clearly there is no unique solution when $r = 1$. The calculations were made by the Rothamsted electronic computer. It was programmed to convert a set of random normal deviates x, taken from Rand Corporation (1955), into a Markov chain y

$$y_i = r y_{i-1} + \sqrt{(1-r^2)}\, x_i, \quad y_1 = x_1$$

$$x_i \text{ is } N(0, 1),$$

and then to estimate m, b and r. 50 sets of 25 x's were used throughout so that, for any given value of n, the y's were derived from the same x's for each value of r and, further, the x's were the first n values from each set of 25. It was not necessary to introduce artificial values for m and b as the solutions obtained in this case would be identical with those for $m = b = 0$, apart from the increase of $\hat{m}$ and $\hat{b}$ by the values used for m and b. The results answer two questions: (1) do direct estimates of mean and regression differ seriously from estimates that recognize the correlation of errors? and (2) is $\hat{r}$ accurate enough to be used as an indicator of the organism's ability to recover from accidental deviations? The answer to the first question is 'no' and to the second, 'not unless n is about 20 or more'.

The mean values of $\hat{m}$ and $\hat{b}$ for each combination of n and r all differed insignificantly from 0, except in the case of $n = 15$ where it is clear that the original data showed a chance disturbance.

Averages of $\hat{m}$ and $\hat{b}$ for $n = 15$

	$100\hat{m}$		$100\hat{b}$	
r	Markov	Direct	Markov	Direct
−0·8	0·87 ± 1·25	1·15 ± 1·43	0·54 ± 0·30	0·96 ± 0·47
−0·4	1·58 ± 2·46	1·61 ± 2·50	1·11 ± 0·56	1·30 ± 0·65
0	2·34 ± 3·70	1·98 ± 3·74	1·71 ± 0·81	1·74 ± 0·83
0·4	2·98 ± 5·46	1·76 ± 5·58	2·44 ± 1·09	2·30 ± 1·08
0·8	−2·59 ± 9·23	−3·33 ± 9·64	3·07 ± 1·32	2·30 ± 1·37

The behaviour of $\hat{m}$ is explained by the consideration that, for $r = 0$, every x has equal importance, whereas as r approaches ± 1 the first x dominates the series. The estimated standard errors for direct estimation, $r = 0$, may be compared with their theoretical values 3·65 and 0·85. It is clear that there is little to choose between the direct and Markov estimates, except perhaps for extreme values of r which would not be encountered in nature. This is hardly surprising as the two kinds of estimate differ only in the weighting of the first and last observations. I am grateful to the Referee for pointing out that, on such theoretical considerations, this weighting would not improve the accuracy of $\hat{m}$ and $\hat{b}$ very much even if an unbiased estimate of r were used. The distributions of $\hat{m}$ and $\hat{b}$ are algebraically quite tractable, but that of $\hat{r}$ is best dealt with numerically.

Means and standard deviations of $\hat{r}$

	$n = 9$		$n = 15$		$n = 15$ (direct)		$n = 20$		$n = 25$	
r	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
−0·8	−0·79	0·18	−0·78	0·17	−0·76	0·17	−0·79	0·15	−0·81	0·12
−0·6	−·64	·26	−·65	·20	—	—	−·65	·18	—	—
−0·4	−·51	·30	−·51	·22	−·50	·21	−·49	·21	−·48	·19
−0·2	−·39	·31	−·36	·23	—	—	−·33	·22	—	—
0	−·27	·33	−·20	·24	−·20	·23	−·16	·23	−·14	·23
0·2	−·16	·33	−·05	·25	—	—	·00	·24	—	—
0·4	−·06	·35	·09	·26	·09	·25	·17	·25	·20	·24
0·6	·02	·36	·22	·26	—	—	·32	·25	—	—
0·8	·06	·37	·33	·27	·30	·24	·47	·24	·51	·23

The 'direct' values of $\hat{r}$ for $n = 15$ were obtained from the residuals after fitting m and b direct, i.e. are the values after the first cycle of iteration. The values of $\hat{r}$ show a negative bias and, more important, a fairly linear regression on r with a coefficient less than 1. Thus the difference in $\hat{r}$ between $r = +0{\cdot}4$ and $-0{\cdot}6$ is 0·82 for $n = 20$ and only 0·58 for $n = 9$. Taking into account the size of the standard deviations, there is little hope of establishing consistent differences between values of r unless n is 20 or more. It appears, therefore, that Markov analysis offers no advance on the simpler methods of estimation of growth parameters (e.g. Leech & Healy, 1959) and a chain of at least 20 observations is required for useful estimates of the error autocorrelation.

References

Leech, F. B. & Healy, M. J. R. (1959). *Biometrics*, **15**, 98.
Quenouille, M. H. (1957). *The Analysis of Multiple Time-Series*. London: Charles Griffin and Co. Ltd.
Rand Corporation (1955). *A Million Random Digits*. Glencoe, Illinois: Free Press.

Corrigendum

Biometrika, (1960), **47**, pp. 202–3.

'Some numerical results for waiting times in the queue $E_k/M/1$.' By C. BURROWS.

p. 202, 2nd line above Table 1

read $\left[\frac{\rho k}{\rho k+1-\lambda}\right]^k$ *for* $\left[\frac{\rho k}{\rho^{k+1}-\lambda}\right]^k$.

Reviews

Probability and Statistics. (The Harald Cramér Volume). Edited by ULF GRENANDER. New York: John Wiley and Sons Inc.; London: Chapman and Hall; Stockholm: Almquist and Wiksell. 1959. Pp. v + 434. 100*s*.

This book is a collection of research papers on probability and statistics presented to Prof. Harald Cramér for his 65th birthday. The titles of the papers and the authors are as follows: Some scaling models and estimation procedures in the latent class model (T. W. Anderson); The impact of stochastic process theory on statistics (M. S. Bartlett); A Markov chain theorem (J. L. Doob); Design of linear experiments (G. Elfving); On combinatorial methods in fluctuation theory (W. Feller); The restricted chi-square test (E. Fix, J. L. Hodges and E. L. Lehmann); Some non-linear problems in probability theory (U. Grenander); Some remarks on stable processes with independent increments (M. Mac); Unitary dilations of Markov transition operators (D. G. Kendall); Construction du processus de W. Feller et H. P. McKean en partant du mouvement Brownien (P. Lévy); Cramér's theorem on monotone matrix-valued functions and the Wold decomposition (P. Masani); Non-linear prediction (P. Masani and N. Wiener); Optimal asymptotic tests of composite statistical hypotheses (J. Neyman); Sequential estimation of the mean of a normal population (H. Robbins); Statistical analysis of stochastic processes with stationary residuals (M. Rosenblatt); A survey of results in the collective theory of risk (C. O. Segerdahl); An introduction to the measurement of spectra (J. W. Tukey); Non-parametric statistical inference (S. S. Wilks); Ends and means in econometric model building (H. Wold).

The majority of the papers are concerned with one aspect or another of stochastic processes, possibly a reflexion on present day preoccupation with that topic. Prof. Cramér has had presented to him an interesting collection of essays and statisticians all over the world will join with the authors in wishing him many happy returns.

F. N. DAVID

Allgemeine Methodenlehre der Statistik. By JOHANN PFANZAGL. (I. Elementäre Methoden unter besonderer Berücksichtigung der Anwendungen in den Wirkschafts- und Sozialwissenschaften) (Band 746/746*a* of Sammlung Göschen). Berlin: Walter de Gruyter and Co. 1960. Pp. 205. 5.80 DM.

This small (4 × 6 in.) handbook contains a concise development of the elementary statistical techniques used in the descriptive treatment of numerical data. The author has managed to include a remarkably comprehensive range of techniques in a very small space, without sacrificing clarity, or adequacy, of treatment. The book is addressed primarily to users of statistics in the fields of economic and social science. For students working in these fields, the book provides a very convenient summary of available elementary statistical methods: the accompanying discussion should even prove adequate for many students learning a modicum of statistics for certain professional examinations.

Chapter headings are as follows:

(1) Fundamental concepts. (2) Frequency distributions. (3) Parameters. (4) Statistical errors. (5) Sampling methods in economic and social statistics. (6) Measures of condition: measures of movement. (7) General theory of measurable characters. (8) The calculation of index numbers. (9) Some examples of index numbers. (10) The analysis of time series. (11) The procurement of numerical data. (12) The computational treatment of numerical data. (13) The representation of numerical data.

The last three chapters are grouped together as a 'Technical Appendix': chapter 13 is, in fact, mainly concerned with diagrammatic representation of data, pictorial examples being provided.

Chapter 7 is of rather special interest. It emphasizes the need for care in deciding upon what is to be measured, and should be a good influence on the work of investigators in economic and social science, if they study this book.

N. L. JOHNSON

Mathematical Statistics. By J. N. KAPUR and H. C. SAXENA. Delhi, Jullundur and Lucknow: S. Chand and Co. 1960. Pp. 400. Rs. 10.

The book is based on the authors' lectures at Delhi University, and it is aimed primarily at University students in the Indian subcontinent. The level of mathematical sophistication is (no doubt deliberately) kept at a very moderate level and the book is chiefly remarkable for the very large number of exercises it contains. The authors claim that 'books on Statistics have so far tended to give either very difficult ...or only very easy exercises, and, in general, these exercises are too few...'. In the present work there are some 600 straightforward exercises, many of them taken from M.A. and M.Sc. papers of Indian universities, many with solutions provided; and in addition there are a large number of worked examples in the text.

This feature is typical of the authors' approach, which, clearly, aims at producing a *teaching* book which will enable even the most plodding of students to pass examinations.

The contents are best illustrated by giving the chapter headings:

Frequency distributions and measures of location; Measures of dispersion, skewness and kurtosis moments; Theory of probability; Mathematical expectation, moment generating and cumulative functions; Special discrete probability distributions; Continuous probability distributions; Special continuous distributions; Method of least squares and curve fitting; Correlation and regression; Multiple and partial correlation; Theory of sampling; Exact sampling distributions; Tests of significance based on t, F and r distributions; Tests of significance based on the chi-square distribution; Statistical theory of estimation; Testing of hypotheses, sequential analysis, distribution-free methods.

In addition there are two appendices, one on beta and gamma functions and one on finite differences. Tables are provided of normal ordinates and areas, percentage points of t and χ^2, and 5 and 1 % values of F.

It is pleasing to see that the Poisson distribution is presented in terms of the Poisson process as well as the usual approximation to the Binomial; and that the definition of independence does not merely refer to the factorization of the joint distribution function, but also makes the necessary remarks about the ranges of definition. On the whole in the greater part of the book, the authors undoubtedly succeed in achieving their aim, keeping the style unpretentious (and often non-rigorous) and keeping the topics discussed very much within 'elementary' bounds.

Those parts of the book dealing specifically with estimation and inference cannot, however, be regarded as entirely satisfactory. The approach to significance tests, exemplified by the following extract, may possibly encourage a rather dogmatic and uncritical attitude in the student: 'If our calculated value of $|t|$ exceeds $t_{0\cdot05}$, we say that the difference between $\bar{x}$ and μ is significant at 5 % level of significance; if it exceeds $t_{0\cdot01}$, the difference is said to be significant at 1 % level or just highly significant. If $|t| < t_{0\cdot05}$, we say that the data is consistent with the hypothesis that μ is the mean of the population' (p. 284). In the chapter dealing with estimation, the probability associated with a confidence interval is described as a fiducial probability and there is no discussion of the differences between the fiducial and the confidence concepts. Finally, the chapter on least squares is very disappointing. There is nothing about any optimum properties of Least Squares estimates, nothing about their accuracy, indeed no indication that Least Squares procedures are in any sense statistical. Frankly this chapter is unworthy of the book.

E. H. LLOYD

Analysis of Straight-line Data. By FORMAN S. ACTON. New York: John Wiley and Sons Inc.; London: Chapman and Hall Ltd. 1959. Pp. xiii + 267. 72*s*.

On the dust-jacket, and perhaps also in the Preface, the author does his book less than justice by saying that it '...thoroughly covers one important aspect of *engineering* statistics...'. The analysis of experimental data which can be associated with the linear model—or for which a linear model provides a reasonable representation—extends beyond the confines of engineering. Indeed, many of the examples are from fields far removed from the world of engineering. Prof. Acton's book is, in fact, a good example of that *developed* stage in the literature of a subject: the monograph. In this case, a statistical concept has been described—no less acceptable to the practising statistician than the monograph which deals with the areas of application for the tools found in one corner of the statistical tool box, or the monograph which deals with the statistical methods appropriate to one field of application.

The author says that his book, grown from a small pamphlet, is a discussion of techniques that have been proved both pertinent and useful. Some of the classical approaches are excluded, however, on the ground that 'they encourage an unthinking mechanical solution to the problem of straight-line fitting'. It is because the experimentalist needs to think about the structure of his experiment that the author makes a strong plea for the experimenter to try and use the methods himself. The only practical difficulty may be that, even with Prof. Acton's book to guide him, the experimenter may be so involved that it could still be preferable to him to consult a statistician rather than face the 250-odd pages of this book and the channels to which they would lead. If this was to be so then it would be a pity because he would miss reading an attractive style of writing. This reviewer is sure that the 'Jacobean' on page 122 is there through a misprint since the author's brand of dry humour would scarcely encompass this on any terms!

The book begins with a short chapter on the choice of (linear) model and then proceeds to two chapters dealing with the classical model—where x is known without error and y homoscedastic—and the regression model where there are several values of y for each known x. In the first of these two chapters, the second section (pages 19–53) dealing with the setting of confidence limits appears to this reviewer to be particularly valuable. Chapters 4 and 5 deal with samples from bivariate normal populations and the problem of regression with both x and y measured with error. This leads to the more general case (in chapter 6) of several regression lines and the use of analysis of variance. Some of the shoulder-headings in these chapters are examples of the author's lively sense that all this is by no means tedious: 'The 99·44 % pure extractor' method and 'Hindsight heresy'.

In chapter 7 the opening paragraph is as distinguished a tilt at the desire of some to pursue the linear form as others which have been written on the hallowed normality of a distribution. The chapter deals with the method of orthogonal polynomials for dealing with data which cannot adequately be represented by a straight line. This, together with the succeeding chapters, (8) on the use of transformations, (9) rejection of unwanted data and (10) on cumulative data, confirms the essentially practical approach of this book. These are short accounts of important problems for which adequate answers have not yet been found.

There is an appendix of useful tables—some from formally unpublished material of Prof. Tukey—the usual index, a list of forty-two quoted references and an Additional Bibliography of annotated references not quoted in the text. In this latter connexion it is good to see included the pre-war paper by Prof. R. G. D. Allen, *Economica* (1939), **6**, 191–204, as indication of the wide experience which has gone into the preparation of this monograph. The diagrams throughout are excellent and it may comfort others to know that even the distinguished house of Wiley is subject to the same kind of printing hazards as most: a figure 1 has dropped-out of the denominator under the root sign in the first term of the first expression on page 133.

This book may not be suitable for students or the recently qualified statistician, but the general body of practising statisticians will do well to note and use this book as a reference point for statistical expertise on its chosen topic.

WM. R. BUCKLAND

Methods of Correlation and Regression Analysis. Third Edition. By M. Ezekiel and K. A. Fox. New York: John Wiley and Sons Inc.; London: Chapman and Hall. 1959. Pp. xv + 548. 88*s*.

The first two editions of this book, written by Prof. Ezekiel, alone appeared in 1930 and 1941 and are familiar to many statisticians. In this third edition the material has been brought up to date by emphasizing regression rather than correlation and by the addition of a new chapter (24) on fitting systems of two or more simultaneous equations. There is also fuller treatment of the analysis of variance. As in previous editions the book sets out how to apply techniques. It is a useful ancillary tool for any research worker who is faced with the problem of studying the relationships between two or more variables.

F. D. DAVID

Modern Statistical Methods. By P. O. Johnson and R. W. B. Jackson. Chicago: Rand McNally and Co. 1959. Pp. viii + 509. $8.00.

This book is designed to meet the needs of students in the social sciences who know no statistical methods and who require to learn enough to be able to treat their practical data. The topics treated are

reduction and classification of data, a little probability, likelihood, χ^2, tests for significance of means and variances, a little analysis of variance, some non-parametric tests, estimation, reduction of bivariate data and regression, multivariate data and partial regression, design and analysis of investigations. Brief tables of the normal integral, of χ^2, of t and of F are included. The book is well and clearly written with many useful numerical illustrations drawn from the field of the social sciences. Care has been taken to give references so that the student who wishes to read further or more deeply about the techniques cited will be able to do so. The book can be recommended.

F. N. DAVID

Introduction to Probability and Statistics. By G. W. McElgrath and B. W. Lindgren. New York and London: Macmillan Company, New York. 1959. Pp. xiii + 273. Price 44*s*.

This is yet another elementary textbook on probability and statistics. There are twelve chapters and a useful appendix containing the more common statistical tables. The authors begin with fives chapters on the elements of probability—Simple models, Random variables and distribution functions, Discreet distributions, Continuous distributions and Sums of random variables. They turn then to the observational tie-up, and give a chapter on sampling and one on the presentation and description of data. There follow chapters on testing statistical hypotheses, testing and estimating location, testing and estimating variability, and two-sample tests. The last chapter touches on sequential tests, regression, analysis of variance and decision theory.

The delineation is clear, if a trifle loose in places, and the book will doubtless serve a useful purpose.

F. N. DAVID

Rapid Statistical Calculations. By M. H. Quenouille. London: Charles Griffin and Co. Ltd. 1959. Pp. xv + 81. 10*s*. 6*d*.

This book is a summary of the tests which are in common statistical practice, and their application to numerical examples. It is divided into three sections: A, Methods involving mean level or scatter of groups of observations. B, Methods of comparing frequencies or proportions. C, Methods involving association. There is an appendix containing brief significance tables for the normal, t, χ^2 and F distributions, for the correlation coefficient, the number of points falling in a quadrant, the number of runs, and the number of pairs of points falling on the same side of the medial line. For students and research workers trying to teach themselves statistics it will be found a useful auxiliary to any elementary theoretical textbook of statistical methods.

F. N. DAVID

Darwin's Biological Work. Some Aspects Reconsidered. Ed. by P. R. Bell. London: Cambridge University Press. 1959. Pp. xiii + 343. 40*s*.

The Darwinian centenary has seen the appearance of a minor spate of books on Darwin, Darwinism and evolution in general. Inevitably some of these were repetitive, others of indifferent scientific merit. Thus the prospective reader may well be sceptical of further Darwiniana. The present volume, however, differs by intent and content from the books referred to above and may well please a wide readership.

The plan of the book is to describe Darwin's contributions to various fields of biology and to review the progress in these fields during the past one hundred years. It is an interesting approach, although not all of the contributors to the volume followed it closely, adding Darwin as an afterthought instead of a *Leitmotif*.

There are six essays in the volume. The first, by the editor, on 'The movement of plants in response to light', is a thorough historical review of the development of our present-day views on the mechanism of the response of plants (mostly seedlings) to light. Approaching this subject from total ignorance I found much informative material, but had some difficulty in co-ordinating all the many facts and

theories. Less citation of factual references and more sectional summaries could possibly have improved its readability.

The article on 'Palaeontology and evolution', by J. Challinor is perhaps the finest in the volume. While it contains little on Darwin as a palaeontologist, it does present a perceptive modern view of the contribution of fossil study to a knowledge of evolution. Dr Challinor is very critical of the palaeontological evidence often cited in support of evolutionary theories by less discerning colleagues. His views of sampling, recognition of trends and classification may not meet with their whole-hearted approval. I found his thoughts on these and other issues stimulating and much to my liking. I am not sure whether the article gains by the description of the fossil evidence in various systematic groups; in any case this discussion has been wisely placed at the end, where also can be found a geological time-scale which will be much appreciated by those, such as this writer, who cannot always recall the entire sequence.

Prof. Haldane, in contributing an article on 'Natural selection', has had the difficult task of condensing this vast subject into 49 pages. The result is a readable and stimulating essay, covering largely familiar ground but with a section on rates of evolution containing novel inferences on the subject based on simple genetic assumptions. Space limitations prevent their discussion here; much of Haldane's reasoning follows along classical selectionist lines. In the opinion of this reviewer insufficient attention is paid to stochastic evolutionary processes, population structure and fluctuation, and to distinguishing catastrophic and selective effects of environmental forces. While surely everyone admires Kettlewell's magnificent study on *Biston betularia*, yet another reproduction of his photographs in this volume is bound to engender *ennui* among the biological public.

P. Marler discusses 'Developments in the study of animal communication'. This copiously illustrated chapter discusses mechanisms of communication with emphasis on the vertebrates and contains fascinating descriptions of various behaviour patterns. The largely original analysis of specific, geographical and individual variation in bird songs will be particularly appreciated.

'Cross: and self-fertilization in plants', by H. L. K. Whitehouse, follows the plan of the volume most admirably and is a lucid, well-organized account of the subject. Darwin's numerous contributions to this field both by experiment as well as theory will greatly impress the reader. The section on the early history of genetics, well known to most, could possibly have been shortened or omitted. On reading through the section on the advantages of heterozygosis one is struck by how little really is known about the relative success of outbred and inbred plants under natural conditions; yet many of the speculations regarding the origins of mechanisms for cross-fertilization hinge upon these facts.

The volume concludes with a scholarly article by J. S. Wilkie on 'Buffon, Lamarck and Darwin: The originality of Darwin's theory of evolution'.

The book is well illustrated. Regrettably, citations to the literature are by number, with all references grouped by chapter at the end of the book, a particularly unhandy system. There is no index.

ROBERT R. SOKAL

Statistical Methods in Biology. By N. T. J. BAILEY. London: English Universities Press. 1959. Pp. ix + 199. 25*s*.

This is a textbook on elementary statistical techniques. The topics covered are the binomial, normal, and Poisson distributions, significance tests (t, χ^2, F and r) contingency, correlation and regression, elements of the analysis of variance and of factorial experiments, partial correlation and regression. No attempt is made (deliberately) to indicate the derivation on any technique. The method of exposition appears to be modelled on that of R. A. Fisher's *Statistical Methods for Research Workers*, but it is unlikely that many will be tempted away from that old and tried classic. The numerical examples illustrating the theory are humdrum. The reviewer for one is tired of the distribution of stature, the distribution of yeast cells in a haemacytometer (Student's data) and the distribution of weights of fathers and sons. Stature and the weights of potatoes can each be described as biological measurements but one of interest to few biologists, and in view of the title it is possibly unfortunate that a greater effort was not made to choose examples which occur in biological work.

This book adds one more to the mass of elementary statistical texts which are now on the market. The time is almost ripe for a monograph which will help the student to choose between them.

F. N. DAVID

String Figures and Other Monographs. By W. W. R. Ball, J. Petersen, H. S. Carslaw and F. Cajori. New York: Chelsea Publishing Co. 1960. Pp. vi + 72, 102, v + 179, vii + 136. $ 3.95.

This is a combined reprinting of four monographs.

The first is W. W. Rouse Ball's 'String Figures'. String figures are designs made by weaving and twisting loops over one's hands; they may either be pleasant in themselves, or illustrate some action or story, or create a surprise effect. Rouse Ball gives examples of such figures from all parts of the world, and references to further literature on the subject.

The second is Petersen's 'Methods and Theories for the Solution of Problems of Geometrical constructions applied to 410 Problems'. This deals with problems soluble by the use of a straight edge and compass. An astonishing variety of problems are reduced in effect to two principles. The first is: 'Imagine one of the given conditions...removed, and seek the loci of the points of the figure thus rendered indeterminate'. The required figure can then be found by means of the intersection of two such loci. The second principle is that of transformation of the figure, by translation, magnification, rotation, inversion, etc., into a new figure whose construction is already known. The author concludes with a few remarks on whether a given problem can be solved with ruler and compasses.

The third monograph is Carslaw's 'The Elements of Non-Euclidean Plane Geometry and Trigonometry'. The scope of this book is well-described by the title. It is very clear and thorough, including both hyperbolic and elliptic geometry, and with full account of the history. Although a modern book might prefer to approach the subject via projective geometry, Carslaw's book is interesting in following more nearly the historical line of development, and is an excellent introduction to and general account of the subject.

The final monograph is Cajori's 'A History of the Logarithmic Slide Rule and allied Instruments'.

The publishers explain that the essential reason for printing all four books in a single volume is economic; it would cost almost the same to print any one of them separately. The result is to get an interesting miscellany at a low price.

CEDRIC A. B. SMITH

Journal of Research. B. Mathematics and Mathematical Physics. Vol. 64B, No. 1. January to March 1960. Washington: U.S. Department of Commerce, National Bureau of Standards. 1960. Pp. 82. Subscription Price: $2.25 (U.S.A.) or $2.75 (elsewhere).

Starting from 1960, this *Journal* is to be published in four separate sections, corresponding to subject matters field. Section B, dealing with Mathematics and Mathematical Physics, will include to$_{\text{l}}$ cs in mathematical statistics, theory of experimental design, numerical analysis, theoretical physics and chemistry, and the logical design and programming of computers.,

The first issue of Section B under the new scheme contains two Selected Bibliographies of Statistical Literature (1930–57) prepared by Lola S. Deming:

I. Correlation and Regression Theory, pp. 55–68.

II. Time series, pp. 69–82.

The titles and references were extracted from a card file, kept in the N. B. S. Statistical Engineering Laboratory, made up of abstracts taken from *Zentralblatt für Mathematik* (1930–39) and from *Mathematical Reviews* (1940–57).

Further bibliographies for other subject groups are promised.

Sampling Inspection Tables. Single and Double Sampling (second edition). By H. F. Dodge and H. G. Romig. London: Chapman and Hall Ltd.; New York: John Wiley and Sons Inc. 1959. Pp. xi + 60 (text), 195 (tables). 64*s*.

This is the second edition revised and expanded of sampling inspection tables first produced in 1944. The additions are:

(i) Use of the tables when rejected lots are not screened.

(ii) Choosing a sampling plan for a specific application.

(iii) A new chapter on the 'operating characteristic curve' (*OC*)

(iv) *OC* curves for all the single and double sampling tables.

The additions have been made at the request of users who have found this book of practical importance.

Other books received

Introduction to Mathematical Statistics. By R. V. Hogg and A. T. Craig. New York and London: The Macmillan Company, NewYork, 1960. Pp. ix + 245.

Type-token Mathematics. By G. Herdan. The Hague, Holland: Mouton and Co. 1960. Pp. 448. Gld. 54 or £5. 2*s*.

Quantative Genetics. By D. S. Falconer. Edinburgh: Oliver and Boyd. 1960. Pp. ix + 365. 35*s*.

Genetical Research. Vol. 1, No. 1. London and New York: Cambridge University Press. 1960. Pp. 172. Subscription (three parts annually) £5 or $17.50.

An Introduction to Infinitely Many Variables. (Griffin's Statistical Monograph No. 6.) By Enders A. Robinson. London: Charles Griffin and Co. 1959. Pp. 132. 28*s*.

An Introduction to Mathematical Statistics. By H. D. Brunk. Massachusetts: Ginn and Company 1960. Pp. xi + 403. $7.00.

Annual Epidemiological and Vital Statistics 1956. (Bilingual edition.) World Health Organization, Geneva, Switzerland. (U.K. Sales Agent, H.M.S.O.) Pp. 705. 60*s*., $12 or Sw. fr. 36.

Essential Tremor. A Clinical and Genetical Population Study. By Tage Larsson and Torsten Sjögren. Copenhagen: Munksgaard, for the Department of Psychiatry, The Caroline Institute, Stockholm. 1960. Pp. 176. Sub. 65 Danish Kronor.

Principles and Procedures of Statistics. By Robert G. D. Steel and James H. Torrie. London, New York and Toronto: McGraw-Hill Publishing Co. 1960. Pp. 481. 81*s*. 6*d*.

Theory and Solution of Ordinary Differential Equations. By Donald Greenspan. New York and London: The Macmillan Company, New York. 1960. Pp. 148. 38*s*. 6*d*.

La Biométrie. Que sais-je. (Le point des connaissances actuelles No. 871.) By Eugène Schreider. Paris: Presses Universitaires de France. 1960. Pp. 126.

The Elliptic Functions as they should be. By Albert Eagle. Cambridge: Galloway and Porter. 1958. Pp. 508. 45*s*.

Biometrical Genetics. Proceedings of an International Symposium, Ottawa 1958, sponsored by the Biometrics Society and the International Union of Biological Sciences [International Union of Biological Science Series B (Colloquia), No. 38 of Monographs on Biometry]. Ed. Oscar Kempthorne. London, Oxford, New York and Paris: Pergamon Press. 1960. Pp. 150. 42*s*.

Biométrie-Praximetrie. Vol. 1, No. 1. Brussels: Presses Universitaires de Bruxelles, for the Association des Biometriciens de Belgique et du Congo Belge Societé Adolphe Quetelet. 1960. Pp. 58. Subscription 500 Belgian Francs p.a. + 5 francs overseas.

Recherche et Expérimentation en Agriculture. Vol. 2. **Methodes Statistiques en Biologie et en Agronomie.** By A. Vessereau. Paris: J. B. Baillière et Fils. 1960. Pp. 538. 55 new French francs.

Modern Mathematics. An Introduction. By Samuel I. Altwerger. New York and London: The Macmillan Company, New York. 1960. Pp. xii + 462. 47*s*.

Special Functions. By Earl D. Rainville. New York and London: The Macmillan Company, New York. 1960. Pp. xii + 365. 82*s*.

Grundlagen der Analysis. By Edmund Landau. New York: Chelsea Publishing Co. 1960. Pp. 173. $1.95.

Prontuari per Calcoli Statistici. Tavole numeriche e complementi. By Silvio Vianelli. Palermo, Italy: Ed. Abbaco. 1959. Pp. 1543. Lire 16,000.

Introduction to Congestion Theory in Telephone Systems. By R. Syski. Edinburgh and London: Oliver and Boyd, for the Automatic Telephone and Electric Co. 1960. Pp. xvi + 742. £5. 5*s*.

Modern Probability Theory and its Applications. By Emanuel Parzen. New York: John Wiley and Sons Inc.; London: Chapman and Hall. 1960. Pp. xv + 464. 86*s*.

Elementary Statistics. Py Paul G. Hoel. New York: John Wiley and Sons Inc.; London: Chapman and Hall. 1960. Pp. vii + 261. 44*s*.

An Introduction to Experimental Design. By William S. Ray. New York and London: The Macmillan Company, New York. 1960. Pp. x + 254. 45*s*. 6*d*.

Probability. An Introduction. By Samuel Goldberg. U.S.A. and London: Prentice Hall Inc. 1960. Pp. xiv + 322. 54*s*.

Statistiche Methoden des Populationsgenetik. By Henri Louis Leroy. Basal and Stuttgart: Birkhäuser Verlag. 1960. Pp. 397. 67.50 Swiss francs.

Studii si Cercetari Matematice. Vol. 11, No. 1. Bucharest: Academia Republicii Populaire Romine. 1960. Pp. 280.

BIOMETRIKA

BACK ISSUES

Volumes 43–47 (1956–1960)

These may be obtained from the BIOMETRIKA OFFICE, at the following prices, which include packing and postage:

£3. 10*s*. (or $10.00) per wrappered volume.
Bound volumes: £1. 8*s*. (or $4.00) extra per volume.
Binding cases: 14*s*. (or $2.00) each.

Cheques should be made payable to *Biometrika*, crossed 'a/c BIOMETRIKA TRUST' and sent to THE SECRETARY, BIOMETRIKA OFFICE, UNIVERSITY COLLEGE LONDON, GOWER STREET, LONDON, W.C. 1, to whom all orders for subscriptions and author's separates should also be sent (see inside front cover). The Secretary also has available for sale copies of many of the papers and tables published in earlier numbers of *Biometrika* (see pages (iv), (v) and (vi)).

Volumes 20A, 20B, 21–42 (1928–1955)

Messrs Wm. Dawson & Sons, Ltd., have been appointed agents for the sale of these volumes, which may be obtained from the address below at the following prices, *packing and postage extra:*

£5. 5*s*. (or $15.00) per wrappered volume.
Bound volumes: £1. 8*s*. (or $4.00) extra per volume.
Binding cases: 14*s*. (or $2.00) each.
Index (Subjects Vols. 1–37; Authors Vols. 1–43): 6*s*. (or $1.00).

Volumes 1–19 (1902–1927)

Messrs Wm. Dawson & Sons, Ltd., have been granted permission to reprint Vols. 1–19. These are ready for distribution, price £130 bound. Single vols. £7. 10*s*. each. Would librarians and others wishing to have copies please place their orders with:

WM. DAWSON & SONS, LTD.,
16 WEST STREET, FARNHAM, SURREY, ENGLAND

(*All rights reserved*)

BIOMETRIKA, Vol. 47, Parts 3 and 4

CONTENTS

PAGE

P. H. LESLIE and J. C. GOWER. The properties of a stochastic model for the predator-prey type of interaction between two species 219
GUNNAR BLOM. Hierarchical birth and death processes. I. Theory 235
GUNNAR BLOM. Hierarchical birth and death processes. II. Applications 245
W. A. GLENN. A comparison of the effectiveness of tournaments 253
S. C. PEARCE. Supplemented balance 263
JOHN LEROY FOLKS and OSCAR KEMPTHORNE. The efficiency of blocking in incomplete block designs 273
FRANK A. HAIGHT. Queueing with balking. II 285
H. A. DAVID and CARMEN A. PEREZ. On comparing different tests of the same hypothesis . . 297
D. J. G. FARLIE. The performance of some correlation coefficients for a general bivariate distribution 307
J. A. MCFADDEN. Two expansions for the quadrivariate normal integral 325
R. G. LAHA and E. LUKACS. On a problem connected with quadratic regression 335
ROGER S. MCCULLOUGH, JOHN GURLAND and LLOYD ROSENBERG. Small sample behaviour of certain tests of the hypothesis of equal means under variance heterogeneity 345
BALKRISHNA V. SUKHATME. Power of some two-sample non-parametric tests 355
W. D. EWAN and K. W. KEMP. Sampling inspection of continuous processes with no autocorrelation between successive results 363
COLIN R. BLYTH and DAVID W. HUTCHINSON. Table of Neyman-shortest unbiased confidence intervals for the binomial parameter 381
B. M. BENNETT and P. HSU. On the power function of the exact test for the 2×2 contingency table 393
J. A. SIMPSON and B. L. WELCH. Table of the bounds of the probability integral when the first four moments are given 399
NORMAN C. SEVERO and MARVIN ZELEN. Normal approximation to the chi-square and non-central F probability functions 411
D. E. BARTON, F. N. DAVID and ANNE F. O'NEILL. Some properties of the distribution of the logarithm of non-central F 417
D. V. LINDLEY, D. A. EAST and P. A. HAMILTON. Tables for making inferences about the variance of a normal distribution 433
D. E. BARTON, F. N. DAVID and M. MERRINGTON. Tables for the solution of the exponential equation, $\exp(-a)+ka = 1$ 439

MISCELLANEA

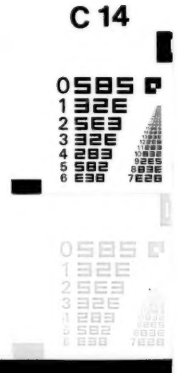
M. G. KENDALL. Studies in the history of probability and statistics. X. Where shall the history of statistics begin? 447
D. R. COX. On the number of renewals in a random interval 449
B. R. BHAT and J. GANI. A note on sufficiency in regular Markov chains 452
R. N. CURNOW. The regression of true value on estimated value 457
W. M. HARPER and J. A. MACDONALD. The use of integral transforms to determine expansions of distribution functions. 460
N. L. JOHNSON and D. H. YOUNG. Some applications of two approximations to the multinomial distribution 463
T. A. RAMASUBBAN. Some distributions arising in the study of generalized differences . . . 469
GEORGE E. BARDWELL. On certain characteristics of some discrete distributions 473
H. LINHART. Approximate tests for m rankings 476
A. M. KSHIRSAGAR. A note on the derivation of some exact multivariate tests 480
N. GILBERT. Estimation from a linear Markov process 482

CORRIGENDUM 484

REVIEWS

Ed. ULF GRENANDER. Probability and Statistics. (The Harald Cramér Volume) 485
JOHANN PFANZAGL. Allgemeine Methodenlehre der Statistik. 485
J. N. KAPUR and H. C. SAXENA. Mathematical Statistics 486
FORMAN S. ACTON. Analysis of Straight-line Data 486
M. EZEKIEL and K. A. FOX. Methods of Correlation and Regression Analysis. Third Edition . 487
P. O. JOHNSON and R. W. B. JACKSON. Modern Statistical Methods 487
G. W. MCELGRATH and B. W. LINDGREN. Introduction to Probability and Statistics . . . 488
M. H. QUENOUILLE. Rapid Statistical Calculations 488
Ed. P. R. BELL. Darwin's Biological Work. Some Aspects Reconsidered 488
N. T. J. BAILEY. Statistical Methods in Biology 489
W. W. R. BALL, J. PETERSEN, H. S. CARSLAW and F. CAJORI. String Figures and Other Monographs 490
Journal of Research. B. Mathematics and Mathematical Physics. Vol. 64B, No. 1 490
H. F. DODGE and H. G. ROMIG. Sampling Inspection Tables. Single and Double Sampling (second edition) 490

OTHER BOOKS RECEIVED 491

Printed in Great Britain at the University Press, Cambridge (Brooke Crutchley, University Printer)

GE
219
235
245
253
263
273
285
297
307
325
335
345
355
363
381
393
399
411
417
433
439
447
449
452
457
460
463
469
473
476
480
482
484
485
485
486
486
487
487
488
488
488
489
490
490
490
491